

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:48:40 ; Search time 2824 Seconds

(without alignments)  
10382.299 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620  
Sequence: 1 gttcaaaataacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 236489745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

GenEmb1:\*  
1: gb\_ba:\*  
2: gb\_hig:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pal:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	14	26.5	165	12 SCU87267	U87267 Synthetic c
2	29	4.7	77612	8 NCB8024	AL669990 Neurospor
3	28	4.5	1845	5 BC044362	BC044362 Danio rer
4	28	4.5	2448	3 AF136344	AF136344 Toxoplasma
5	28	4.5	204907	10 AL772249	AL772249 Mouse DNA
6	27	4.4	251	6 BD213305	BD213305 Novel hum
7	27	4.4	927	3 AMTRVPLP1	AMTRVPLP1
8	27	4.4	966	3 BC004856	BC004856 Homo sapi
9	27	4.4	966	3 BC004856	BC004856 Homo sapi
10	27	4.4	1453	9 AY130010	AY130010 Homo sapi
11	27	4.4	1629	10 BC049168	BC049168 Mus muscu
12	27	4.4	1705	5 BC044682	BC044682 Xenopus l
13	27	4.4	1992	5 BC044682	BC044682 Drosophila
14	27	4.4	2636	3 AY069866	AY069866 Leishmani
15	27	4.4	2905	3 AY099299	AY099299 Leishmani
16	27	4.4	2916	3 AY099298	AY099298 Leishmani
17	27	4.4	4548	3 AF414118	AF414118 Drosophila
18	27	4.4	101002	3 AY216936	AY216936 Plasmodi
19	27	4.4	101016	3 AY216939	AY216939 Plasmodi

c 20	27	4.4	101095	3 AY216937	AY216937 plasmodi
c 21	27	4.4	101158	3 AY216938	AY216938 plasmodi
c 22	27	4.4	160310	5 BX511089	BX511089 zebrafish
c 23	27	4.4	165777	10 BX004852	BX004852 mouse DNA
c 24	27	4.4	198057	2 AC073719	AC073719 Mus muscu
c 25	27	4.4	199866	2 AC073872	AC073872 Plasmodi
c 26	27	4.4	231443	3 AC073691	AC073691 Mus muscu
c 27	27	4.4	231524	2 AC095074	AC095074 Rattus no
c 28	27	4.4	253358	2 AC138303	AC138303 Mus muscu
c 29	26	4.2	157	6 CO466276	CO466276 Sequence
c 30	26	4.2	383	6 CO427072	CO427072 Sequence
c 31	26	4.2	614	3 AF299080	AF299080 Schistos
c 32	26	4.2	694	10 BC049612	BC049612 Mus muscu
c 33	26	4.2	793	6 BD019673	BD019673 Novel gen
c 34	26	4.2	793	6 BD098611	BD098611 Novel gen
c 35	26	4.2	993	6 AB035262	AB035262 Bipolaris
c 36	26	4.2	1054	6 AX214454	AX214454 Sequence
c 37	26	4.2	1054	6 AX951565	AX951565 Sequence
c 38	26	4.2	1185	3 AY094632	AY094632 Drosophila
c 39	26	4.2	1384	6 BD192727	BD192727 123 human
c 40	26	4.2	1384	6 AR184198	AR184198 Sequence
c 41	26	4.2	1696	6 AX685169	AX685169 Sequence
c 42	26	4.2	2066	3 AB072305	AB072305 Bombyx mo
c 43	26	4.2	2081	9 AB056423	AB056423 Macaca fa
c 44	26	4.2	2625	3 AY078995	AY078995 Brachios
c 45	26	4.2	2628	3 AY095026	AY095026 Drosophila

## ALIGNMENTS

RESULT 1	SCU87267	165 bp	RNA	linear	SYN 05-MAR-1997
LOCUS	SCU87267				
DEFINITION	Synthetic construct promoting cardiac myofibrillogenesis in				
ACCESSION	U87267				
VERSION	U87267.1	GI:1842072			
KEYWORDS					
SOURCE	synthetic construct				
ORGANISM	artificial sequences.				
REFERENCE	1 (bases 1 to 165)				
AUTHORS	Lemanski, L.F., Nakatogawa, M., Bhatia, R., Erginel-Unaltuna, N., Spinner, B., and Dude, D.K.				
TITLE	A specific synthetic RNA promotes cardiac myofibrillogenesis in the Mexican axolotl				
JOURNAL	Biochem. Biophys. Res. Commun. 229 (3), 974-981 (1996)				
MEDLINE	97115686				
PUBMED	8955002				
REFERENCE	2 (bases 1 to 165)				
AUTHORS	Lemanski, L.F., Nakatogawa, M., Bhatia, R., Erginel-Unaltuna, N., Spinner, B., and Dude, D.K.				
TITLE	Direct Substitution				
JOURNAL	Submitted (27-JAN-1997) Anatomy and Cell Biology, SUNY-Health Science Center, 750 East Adams Street, Syracuse, NY 13210, USA				
FEATURES	Location/Qualifiers				
source	1..165				
	/organism="synthetic construct"				
	/mol_type="genomic RNA"				
	/db_xref="taxon:32630"				
	/lab_host="Ambystoma mexicanum"				
	/note="the synthetic construct was prepared using a sequence isolated from an axolotl cDNA library lab host strain=Lemanski"				
ORIGIN					
Query Match	26.5%; Score 164; DB 12; Length 165;				
Best Local Similarity	100.0%; Pred. No. 7.2e-87;				
Matches	164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
312	AGACACATTCATTTTGGACACCTCTCTACCGTGAGAGAGGAGCCGATCTT	371			



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 16, 2004, 13:21:52 / Search time 167.5 Seconds

(without alignments)  
4259.496 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 201  
Sequence: 1 gttcaaaataacatttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1825181 seqs, 575374646 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3650028

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame\_n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool.P/US10822496/runac.16112004.100156.24597/app.query.fasta\_1.775  
-DB=uniprot\_02 -QFMT=fasta -SUFFIX=colindp.rup -MINMATCH=0.1 -LOOPT=0  
-LOOEXT=0 -UNITS=bits -SPART=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10822496 -CGN 1 1 244 -runac.16112004.100156.24597 -NCP=6 -ICPU=3  
-NO MAP -LANG=QUEY -NEG SCORES=0 -WAIT -DSPELLOC=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6  
-Fgapext=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Uniprot\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_crembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	9	4.4	45	Q7RM43	Q7RM43 plasmodium
C 2	9	4.4	76	Q6Z201	Q6Z201 pyrodicticum
C 3	9	4.4	111	Y62C_YEAST	P53245 saccharomyc
C 4	9	4.5	120	Q7RLY6	Q7RLY6 plasmodium
C 5	9	4.5	130	Q9BJX1	Q9BJX1 plasmodium
C 6	9	4.4	150	Q95WU0	Q95WU0 trypanosoma
C 7	9	4.4	196	Q81JEO	Q81JEO plasmodium
C 8	9	4.4	198	Q94DG3	Q94DG3 cryza sativ
C 9	9	4.5	214	Q816Z8	Q816Z8 plasmodium
C 10	9	4.5	221	Q7RID7	Q7RID7 plasmodium
C 11	9	4.5	240	Q6Q143	Q6Q143 rictus norv
C 12	9	4.5	240	Q6Q143	Q6Q143 rictus norv
C 13	9	4.5	242	Q6P667	Q6P667 homo sapien
C 14	9	4.5	242	AAH62433	AAH62433 homo sapi
C 15	9	4.5	246	Q8CFP4	Q8CFP4 mus musculi
C 16	9	4.4	259	Q6JCU2	Q6JCU2 aleurodicus

17	9	4.5	280	Q9UF68	Q9UF68 homo sapien
18	9	4.5	294	Q9AKW7	Q9AKW7 cryza sativ
C 19	9	4.4	294	Q8RPI2	Q8RPI2 battonella
20	9	4.5	296	Q9WUW2	Q9WUW2 mus musculi
21	9	4.5	303	Q6PFA7	Q6PFA7 mus musculi
22	9	4.5	303	AAH57655	AAH57655 mus muscu
23	9	4.5	358	Q8XV92	Q8XV92 ralatonia s
24	9	4.5	381	Q9H8P5	Q9H8P5 homo sapien
25	9	4.5	381	Q9H8P5	Q9H8P5 homo sapien
26	9	4.5	381	Q80UY2	Q80UY2 mus musculi
C 27	9	4.4	409	Q6PNA8	Q6PNA8 steinernema
C 28	9	4.4	409	Q6PNA8	Q6PNA8 steinernema
29	9	4.5	454	Q7RFH7	Q7RFH7 plasmodium
30	9	4.5	645	Q7R6E8	Q7R6E8 plasmodium
31	9	4.5	677	Q7R9G4	Q7R9G4 plasmodium
32	9	4.5	765	Q7RIN2	Q7RIN2 plasmodium
33	9	4.5	784	GCF_HUMAN	GCF_HUMAN
34	9	4.5	825	Q81ZK4	Q81ZK4 plasmodium
C 35	9	4.4	929	Q7RJ80	Q7RJ80 plasmodium
C 36	9	4.5	957	Q7RANH	Q7RANH plasmodium
37	9	4.5	1064	Q7RCM9	Q7RCM9 plasmodium
38	9	4.5	2010	Q96171	Q96171 plasmodium
39	9	4.5	2065	Q7RHY8	Q7RHY8 plasmodium
40	9	4.5	3604	Q7RRF7	Q7RRF7 plasmodium
41	9	4.5	4230	Q81C31	Q81C31 plasmodium
42	9	4.5	8094	Q81LB9	Q81LB9 plasmodium
C 43	8	3.9	18	Q8HKH6	Q8HKH6 apnomoma un
C 44	8	3.9	38	Q7RDE8	Q7RDE8 plasmodium
C 45	8	3.9	43	Q9HBA8	Q9HBA8 homo sapien

#### ALIGNMENTS

RESULT 1  
ID Q7RM43 PRELIMINARY; PRT; 45 AA.  
AC Q7RM43;  
DT 01-MAR-2004 (TEMBLrel. 26, Created)  
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)  
DE Hypothetical protein.  
GN Name=PY02345;  
OS Plasmodium yoelii yoelii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=73239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17XN1;  
RX PubMed=1236865;  
RA Carlton U.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perrea M.,  
RA Silva J.C., Ermolaeva W.D., Allen J.E., Selengut J.D., Koo H.L.,  
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
RA Snallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
RA Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,  
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
RA van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,  
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
RA Carucci D.V.I.,  
RT "Genome sequence and comparative analysis of the model rodent malaria  
RT parasite Plasmodium yoelii yoelii.",  
RL Nature 419:512-519 (2002).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL: AABL0100640; EAA21783.1; -  
KM HYPOHETICAL PROTEIN  
SQ SEQUENCE 45 AA, 5330 MM, 7C9B9638F4C9EAD77 CRC64;  
Alignment Scores: 13.1 Length: 45  
Pred. No.: 9.00 Matches: 9  
Score: 100.00% Conservative: 0

US-10-822-496-5	(1-620)	x	Q7RM43	(1-45)
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	4.41%	Indels:	0	
DB:	2	Gaps:	0	

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 16, 2004, 13:47:29 ; Search time 122.5 Seconds  
(without alignments)  
3581.516 Million cell updates/sec

Title: US-10-822-496-5  
Perfect score: 201  
Sequence: 1 gttcaaaataacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1568699 seqs, 353819137 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3023662

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model -DEV=xlp  
-O/cgna2.1/uspro.spcol.p/US10822496/runat.16112004.100159.24707/app.query.fasta.1.775  
-DB=Published Applications AA -QWMT=faetan -SUFPRX=olin2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITG=bits -STAR=1 -END=1 -PATRX=olin2p.rapb -MINMATCH=0.1  
-TRAN=human40.coi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIDE=500 -MINLEN=0  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-Fgapop=6 -Fgapext=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: Published Applications AA:

1: /cgna2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgna2\_6/ptodata/1/pubpaa/PTC\_NEW\_PUB.pep.\*  
3: /cgna2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgna2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgna2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgna2\_6/ptodata/1/pubpaa/PTCUS\_PUBCOMB.pep.\*  
7: /cgna2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgna2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgna2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgna2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgna2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgna2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep.\*  
13: /cgna2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgna2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgna2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgna2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgna2\_6/ptodata/1/pubpaa/US10E\_NEW\_PUB.pep.\*  
18: /cgna2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgna2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgna2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	10	4.9	87	17	US-10-425-115-280203	Sequence 280203,
2	10	5.0	656	17	US-10-739-930-5580	Sequence 5580, Ap
3	9	4.5	17	10	US-09-876-904A-345	Sequence 345, Ap
4	9	4.4	31	9	US-09-989-919-112	Sequence 112, App
5	9	4.5	37	17	US-10-425-115-208638	Sequence 208638,
6	9	4.4	39	9	US-09-925-299-1515	Sequence 1515, Ap
7	9	4.4	39	10	US-09-925-299-1515	Sequence 1515, Ap
8	9	4.4	40	17	US-10-425-115-354630	Sequence 354630,
9	9	4.5	41	16	US-10-437-963-174933	Sequence 174933,
10	9	4.4	41	17	US-10-425-115-238266	Sequence 238266,
11	9	4.4	42	17	US-10-425-115-309326	Sequence 309326,
12	9	4.4	44	15	US-10-264-049-2354	Sequence 2354, Ap
13	9	4.4	44	15	US-10-424-599-152977	Sequence 152977,
14	9	4.4	44	15	US-10-424-599-213738	Sequence 213738,
15	9	4.4	47	16	US-10-437-963-163698	Sequence 163698,
16	9	4.4	47	17	US-10-425-115-274699	Sequence 274699,
17	9	4.4	50	16	US-10-437-963-182797	Sequence 182797,
18	9	4.4	51	9	US-09-764-846-151	Sequence 151, App
19	9	4.4	51	14	US-10-091-483-151	Sequence 151, App
20	9	4.4	51	17	US-10-425-115-229065	Sequence 229065,
21	9	4.4	52	15	US-10-424-599-166357	Sequence 166357,
22	9	4.4	52	17	US-10-425-115-352124	Sequence 352124,
23	9	4.4	53	15	US-10-424-599-274416	Sequence 274416,
24	9	4.4	54	16	US-10-767-701-50915	Sequence 50915, A
25	9	4.4	54	17	US-10-425-115-319798	Sequence 319798,
26	9	4.5	57	17	US-10-425-115-289672	Sequence 289672,
27	9	4.4	58	14	US-09-764-846-183	Sequence 183, App
28	9	4.4	58	17	US-10-091-483-183	Sequence 183, App
29	9	4.4	58	17	US-10-425-115-208769	Sequence 208769,
30	9	4.4	59	17	US-10-425-115-208268	Sequence 208268,
31	9	4.4	59	17	US-10-425-115-269560	Sequence 269560,
32	9	4.4	60	15	US-10-424-599-198167	Sequence 198167,
33	9	4.4	60	17	US-10-425-115-265099	Sequence 265099,
34	9	4.4	61	17	US-10-425-115-330384	Sequence 330384,
35	9	4.4	61	9	US-09-764-846-195	Sequence 195, App
36	9	4.4	61	14	US-10-091-483-195	Sequence 195, App
37	9	4.4	61	15	US-10-424-599-239354	Sequence 239354,
38	9	4.4	62	17	US-10-425-115-337056	Sequence 337056,
39	9	4.4	63	15	US-10-424-599-149489	Sequence 149489,
40	9	4.4	63	15	US-10-424-599-178837	Sequence 178837,
41	9	4.4	63	17	US-10-425-115-335140	Sequence 335140,
42	9	4.4	63	17	US-10-425-115-368838	Sequence 368838,
43	9	4.4	64	15	US-10-424-599-177321	Sequence 177321,
44	9	4.5	64	17	US-10-425-115-218215	Sequence 218215,
45	9	4.5	64	17	US-10-425-115-356691	Sequence 356691,

## ALIGNMENTS

RESULT 1  
US-10-425-115-280203  
; Sequence 280203, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: la Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21 (3322)B  
; CURRENT APPLICATION NUMBER: US/10/425, 115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 280203  
; LENGTH: 87  
; TYPE: PRT  
; ORGANISM: Zea mays  
; OTHER INFORMATION: Clone ID: MRT4577\_187121C.1.pep  
; US-10-425-115-280203



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 16, 2004, 13:33:28 ; Search time 28.5 Seconds

(without alignments)  
2885.415 Million cell updates/sec

```
Title: US-10-822-496-5
Perfect score: 201
Sequence: 1 gtccaaataacatttcaa.....aaaaaaaaaaaaaaaaaa 620
```

Xgapop	60.0	Xgapext	60.0
Ygapop	60.0	Ygapext	60.0
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 478139 seqs, 66318000 residues

Word size: 1

Total number of hits satisfying chosen parameters: 839080

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

### Command line parameters

```

MODEL=1;time=25;model=DBSCAN;
Q=CG12.1/US10;spool=/US10822496/runat_16112004.100157_24623/app_query.fasca.1.7759
DB=Issued/Patents_AA-QWAT=fastan-SUFFIX=olimp;rai-PLANNMATCH=0.1-LOOPPCL=0
LOOPEXT=0-UNITS=bits-START=1-END=1-MATX=olimp;oligo-TRANS=human40.0.cdi
LIST=45-DOCLINK=100-HEAPSIZE=500-MINLEN=0-MAXLEN=2000000000
OUTPM=pcio-NORM=ext-HEAPSIZE=500-MINLEN=0-MAXLEN=2000000000
USFR=US10822496@CCN1.1.46@runat_16112004.100157_24623-NCPI=6-ICPU=3
NO_WMAP-LARGEQUERY-NEG_SCORES=0-WAIT-DSBLOG=100-NOLOG
DELT_TIMEOUT=120-WARN_TIMEOUT=30-THRAADS=1-XGAPOP=60-FGAPOP=6
FGAEXT=7-XGAPOP=60-LOOPEXT=60-DELT=7

```

Database : Issued\_Patents\_AA:\*

```

1: /cgn2_6/ptodata/1/1aa/5A_CONB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_CONB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_CONB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_CONB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCTUS_CONB.pep:*
6: /cgn2_6/ptodata/1/1aa/backtills1.pep:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9	4.5	94	4	US-09-248-796A-22230	Sequence 22230, App
2	9	4.5	285	4	US-09-270-767-44439	Sequence 44439, App
3	8	4.0	13	4	US-09-352-601-11	Sequence 11, App
4	8	4.0	15	4	US-09-082-358B-24	Sequence 24, App
5	8	4.0	16	3	US-08-851-843A-201	Sequence 201, App
6	8	4.0	16	3	US-08-974-549A-320	Sequence 320, App
7	8	4.0	16	3	US-08-854-050-201	Sequence 201, App
8	8	4.0	16	3	US-09-430-323-201	Sequence 201, App
9	8	4.0	16	4	US-09-402-181B-320	Sequence 320, App
10	8	4.0	16	4	US-09-721-456-320	Sequence 320, App
11	8	3.9	17	4	US-09-146-545-221	Sequence 221, App
12	8	4.0	35	4	US-09-513-599C-6560	Sequence 6560, App

C	45	8	3.9	505	4	US-09-270-767-64185	Sequence 4315, A
C	44	8	4.0	480	3	US-09-100-193-1	Sequence 1, A
C	43	8	3.9	438	4	US-09-248-796A-19853	Sequence 19853, A
C	42	8	4.0	438	4	US-09-270-767-44166	Sequence 20799, A
C	41	8	4.0	438	4	US-09-252-991A-20799	Sequence 41669, A
C	40	8	4.0	403	4	US-09-248-796A-19438	Sequence 19438, A
C	39	8	3.9	391	4	US-09-248-796A-15698	Sequence 15698, A
C	38	8	4.0	375	4	US-09-328-352-4967	Sequence 4967, A
C	37	8	4.0	345	4	US-09-248-796A-20831	Sequence 20831, A
C	36	8	4.0	339	2	US-08-855-714-3	Sequence 3, A
C	35	8	4.0	335	4	US-09-252-991A-26119	Sequence 26119, A
C	34	8	4.0	332	3	US-08-178-423A-2	Sequence 2, A
C	33	8	4.0	332	3	US-08-178-423A-2	Sequence 4, A
C	32	8	4.0	326	2	US-08-538-642-2	Sequence 2, A
C	31	8	4.0	246	4	US-09-338-092-167	Sequence 167, A
C	30	8	4.0	215	4	US-09-270-767-44502	Sequence 44502, A
C	29	8	4.0	182	4	US-09-489-039A-8236	Sequence 8236, A
C	28	8	4.0	178	4	US-09-489-039A-8236	Sequence 8236, A
C	27	8	3.9	166	4	US-09-543-681A-8310	Sequence 8310, A
C	26	8	4.0	146	4	US-09-248-796A-11159	Sequence 11159, A
C	25	8	3.9	143	4	US-09-270-767-51028	Sequence 51028, A
C	24	8	3.9	143	4	US-09-270-767-35811	Sequence 35811, A
C	23	8	3.9	139	4	US-09-270-767-55283	Sequence 56283, A
C	22	8	4.0	139	4	US-09-270-767-41067	Sequence 41067, A
C	21	8	3.9	112	4	US-09-543-681A-8315	Sequence 8315, A
C	20	8	3.9	79	4	US-09-248-796A-27552	Sequence 27552, A
C	19	8	4.0	74	4	US-09-621-976-5184	Sequence 5182, A
C	18	8	4.0	71	4	US-09-621-976-5184	Sequence 5184, A
C	17	8	4.0	67	4	US-09-248-796A-74256	Sequence 74256, A
C	16	8	3.9	67	4	US-09-270-767-44621	Sequence 44621, A
C	15	8	3.9	66	4	US-09-248-796A-45981	Sequence 45981, A
C	14	8	4.0	60	4	US-09-248-796A-24364	Sequence 24364, A
C	13	8	4.0	59	4	US-09-621-976-5183	Sequence 5183, A
C	12	8	4.0	59	4	US-09-621-976-5183	Sequence 5183, A

## ALIGNMENTS

## RESULT 1

US-09-248-796A-22230  
; Sequence 22230, Application US/09248796A  
; Patent No. 6747137

```

; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al

```

1. TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A

```

;  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,722

;  
PRIOR FILING DATE: 1998-02-13  
;  
PRIOR APPLICATION NUMBER: US 60/096,400

```

; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208

```

```

; SEQ ID NO 22230
; LENGTH: 94

```

```

; TYPE: PRT
; ORGANISM: Candida albicans

```

US-09-248-796A-22230

Alignment Scores:  
Pred. No.: 2.17

Score:	9.00
Percent Similarity:	100.00

```
Best Local Similarity: 100.00%
Query Match: 4.48%
```

DE: 4

US-10-822-496-5 (1-620) X US-0

594 AACGAGAAATAAAAAA

DB 16 ASNGIULYSLSLSLS



RESULT 2  
US-09-270-767-44439

```

Sequence 44439, Application US/09270767
; Patent No. 6703491
;
; GENERAL INFORMATION:
;
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ. ID NOS: 62517
;
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44439
;
; LENGTH: 285
;
; TYPE: PR1
;
; ORGANISM: Drosophila melanogaster
;
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
;
US-09-270-767-44439

```

Alignment Scores:

Pred. No.:	1.88	285
Score:	9.00	
Percent Similarity:	100.00%	Matches: 9
Best Local Similarity:	100.00%	Conservative: 0
Query Match:	4.48%	Mismatches: 0
	4	Indels: 0
DB:		Gaps: 0

US-10-822-496-5 (1-620) X US-09-270-767-44439 (1-285)

594 AACGAGAAAAAATAAAAAA 620

Db 190 AsnglLysLysLysLysLys 196

### RESULT 3

```

US-09-325-601-11
Sequence 11, Application US/09325601
Patent No. 6573045
GENERAL INFORMATION:
APPLICANT: Karn
APPLICANT: Prescott
TITLE OF INVENTION: Methods and kits for Discovery of RNA-Binding Compounds
FILE REFERENCE: 1950/81235
CURRENT APPLICATION NUMBER: US/09/325,601
CURRENT FILING DATE: 1999-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 13
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence:Peptide from K,
OTHER INFORMATION: Q, R-pentapeptide library
US-09-325-601-11

```

Alignment Scores:

Pred. No.:	26.9	Length:	8
Score:	8.00	Matches:	1
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.98%	Indels:	0
DB:	4	Gaps:	0

US-10-822-496-5 (1-620) X US-09-325-601-11 (1-13)

QY 595 ACGAGAAAAA 618

Db 1 ThrArgGLysLysLysLysLys 8

## RESULT & DISCUSSION

US-09-082-358B-24  
; Sequence 24, Application US/09082358B

; Patent No. 6469153  
; GENERAL INFORMATION  
; ADDITIONAL OFFICE

```

1  APPLICANT: Geoff, Stephen P.
2  APPLICANT: Li, Xiangming
3  TITLE OF INVENTION: EIP-1, EIP-3 GENES, ENVELOPE-INTERACTING PROTEINS
4  TITLE OF INVENTION: EIP-1, and EIP-3
5  FILE REFERENCE: 0575/54804
6  CURRENT APPLICATION NUMBER: US/09/082,356B
7  CURRENT FILING DATE: 1998-05-20
8  NUMBER OF SEQ ID NOS: 106
9  SOFTWARE: PatentIn Ver. 2.1
10 SEQ ID NO 24
11 LENGTH: 15
12 TYPE: prt
13 ORGANISM: murine
14 US-09-082-356B-24

```

Alignment Scores:  
Pred No.

Pred. No.:	26.4	Length:	15
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.98%	Indels:	0
DB:	4	Gaps:	0

US-10-822-496-5 (1-620) x US-09-082-358B-24 (1-15)

[illegible]

Db 5 GluLysLysLysLysLysLys 12

RESULT 5  
ITS-08-85

US-08-851-843A-201  
 / Sequence 001, Application US/08851843A  
 / Patent No. 6093809  
 GENERAL INFORMATION:  
 APPLICANT: Cech, Thomas R.  
 APPLICANT: Lingner, Joachim  
 APPLICANT: Nakamura, Toru  
 APPLICANT: Chapman, Karen B.  
 APPLICANT: Morin, Gregg B.  
 APPLICANT: Harley, Calvin  
 APPLICANT: Andrews, William H.  
 TITLE OF INVENTION: No. 6093809el Telomerase  
 NUMBER OF SEQUENCES: 225  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, 8th Floor  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/851,843A  
 FILING DATE: 06-MAY-1997  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/846,017  
 FILING DATE: 25-APR-1997  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/844,419  
 FILING DATE: 18-APR-1997  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/724,643  
 FILING DATE: 01-OCT-1996  
 CLASSIFICATION:



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 16, 2004, 13:20:52 ; Search time 125.5 Seconds

(without alignments)  
3544.417 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 201  
Sequence: 1 gtcacaaataacatttta.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2002273 seqs, 358729299 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3770738

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool.p/US10822496/runat.16112004.100155.24591/app.query.fasta\_1.775  
-DB=A.Geneseq.23Sep04 -QEMT=fastan -SUFFIX=colin2p.rag -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdt  
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10822496\_QCGN\_1\_1\_224\_@runat.16112004.100155.24591 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6  
-Fgapext=7 -Ygapop=60 -Ygapext=60 -DELop=6 -DELext=7

Database :

A.Geneseq.23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	10	4.9	95	AA009153	
C 2	10	4.9	121	AA005271	AA009153 Human pol
C 3	10	4.9	122	AA002187	AA005271 Human pol
C 4	9	4.5	17	AB074581	AA002187 Human pol
C 5	9	4.4	25	AA008699	AB074581 Transcript
C 6	9	4.4	31	AB074338	AA008699 Human pol
C 7	9	4.4	39	AA011858	AB074338 Human pol
C 8	9	4.4	43	AA011858	AA011858 Human col
C 9	9	4.4	44	ABP41222	AA011858 Human col
C 10	9	4.4	46	AA007812	ABP41222 Human cva
					AA007812 Human pol

C 11	9	4.4	49	AA004514	AA004514 Human pol
C 12	9	4.4	51	AA005079	AA005079 Human pol
C 13	9	4.4	51	AA013671	AA013671 Human pol
C 14	9	4.4	51	AA018166	AA018166 Novel hum
C 15	9	4.4	51	AB092587	AB092587 Human DNA
C 16	9	4.4	51	AD025304	AD025304 Human ext
C 17	9	4.5	57	AA011486	AA011486 Human pol
C 18	9	4.4	58	AA013804	AA013804 Human pol
C 19	9	4.4	58	AA018198	AA018198 Human pol
C 20	9	4.4	58	AB092619	AB092619 Human DNA
C 21	9	4.4	61	AD025316	AD025316 Human ext
C 22	9	4.4	61	AA018210	AA018210 Novel hum
C 23	9	4.4	61	AB092631	AB092631 Human DNA
C 24	9	4.4	61	AD025348	AD025348 Human ext
C 25	9	4.4	71	AA005367	AA005367 Human pol
C 26	9	4.4	78	AA002152	AA002152 Human pol
C 27	9	4.5	84	ABR01760	ABR01760 Human bre
C 28	9	4.4	87	AA013788	AA013788 Human pol
C 29	9	4.4	88	AA003281	AA003281 Human pol
C 30	9	4.4	89	AA007067	AA007067 Human pol
C 31	9	4.4	91	AA006581	AA006581 Human pol
C 32	9	4.4	91	ABP10718	ABP10718 Human ORF
C 33	9	4.4	102	AA007380	AA007380 Human pol
C 34	9	4.4	102	AA010019	AA010019 Human pol
C 35	9	4.4	103	AA005593	AA005593 Human pol
C 36	9	4.4	104	AA007033	AA007033 Human pol
C 37	9	4.4	106	AA007378	AA007378 Human pol
C 38	9	4.4	106	AA018225	AA018225 Plasmid
C 39	9	4.4	106	AA007378	AA007378 Human pol
C 40	9	4.4	106	AB023529	AB023529 Plasmid
C 41	9	4.4	111	AA002315	AA002315 Human pol
C 42	9	4.4	111	AA011818	AA011818 Human pol
C 43	9	4.4	111	AA004871	AA004871 Human pol
C 44	9	4.4	112	AB093107	AB093107 S. cerevi
C 45	9	4.4	114	AA003128	AA003128 Human pol
					AA031177 Novel hum

#### ALIGNMENTS

RESULT 1	AA009153	AA009153 standard; protein, 95 AA.
AC	AA009153;	
XX		
DT	06-NOV-2001 (first entry)	
XX		
DE	Human polypeptide SEQ ID NO 23045.	
XX		
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vacuole; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	nervous system disorders; arthritis; inflammation.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200164835-A2.	
XX		
PD	07-SEP-2001.	
XX		
PF	26-FEB-2001; 2001WO-US004927.	
XX		
PR	28-FEB-2000; 2000US-00515126.	
XX		
PR	18-MAY-2000; 2000US-00577409.	
XX		
PA	(HYSE) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Drmanac RT;	
XX		
DR	WPI; 2001-514838/56.	
XX		
DR	N-PSDB; AA189084.	
XX		
PT	Isolated nucleic acids and polypeptides, useful for preventing diagnosing	

PT	and treating e.g. leukemia, inflammation and immune disorders.
XX	
PS	Claim 20; SEQ ID NO 23045; 1399pp + Sequence Listing; English.
XX	
CC	The invention relates to human polynucleotides (AA179941-AA193841) and
CC	the encoded proteins (AA000010-AA013910), that exhibit activity elating to
CC	cyclokin, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cyclokin-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation. Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_poc_sequences
XX	
SQ	Sequence 95 AA;
Alignment Scores:	
Pred. No.:	2.65 Length: 95
Score:	10.00 Matches: 10
Percent Similarity:	100.00% Conservatve: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	4.90% Indels: 0
DB:	Gaps: 0
US-10-822-496-5 (1-620) x AA009153 (1-95)	
OY	620 TTTTNTTTTTTTTTTTTTTTTCGTTGGA 591
Db	 78 PhePhePhePhePhePheValArg 87
RESULT 2	
AA005271	
ID	AA005271 standard; protein; 121 AA.
XX	
XX	AA005271;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 19163.
XX	
KM	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoesis;
KW	tissue growth factor; immunomodulatory; cancer; leukemia;
KW	nervous system disorders; arthritis; inflammation.
XX	
OS	Homo sapiens.
XX	
PN	WO200164835-A2.
PD	
XX	07-SEP-2001.
XX	
PF	26-FEB-2001; 2001WO-US004927.
XX	
PR	28-FEB-2000; 2000US-00515126.
ER	18-MAY-2000; 2000US-00577409.
XX	
FA	(HYSB-) HXSEO INC.
XX	
PI	Tang YT, Liu C, Dymnac RT;
XX	
WI	WPI; 2001-514838/56.
DR	N-Psdb; AAI85202.
XX	
PT	Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX	and treating e.g. leukemia, inflammation and immune disorders.
PS	Claim 20; SEQ ID NO 19163; 1399pp + Sequence Listing; English.
XX	
CC	The invention relates to human polynucleotides (AA179941-AA193841) and

	SQ	Sequence 121 AA;			
	Alignment Scores:				
Cy	Pred. No.:	2.55	Length:	121	
	Score:	10.00	Matches:	10	
	Percent Similarity:	100.00%	Conservative:	0	
	Best Local Similarity:	100.00%	Mismatches:	0	
	Query Match:	4.90%	Indels:	0	
DB:	Gaps:	4	Gaps:	0	
US-10-822-496-5 (1-620) x AAO05271 (1-121)					
Dc	620 TTTTCTTTTTTTTTTTTCGCGTCCA	591			
Dd	101 PhephephephephepheleuValArg	110			
RESULT 3					
AAO02187	ID	AAO02187 standard; protein; 122 AA.			
AC	AAO02187;				
DT	06-NOV-2001 (first entry)				
DE	Human polypeptide SEQ ID NO 16079.				
KM	Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; hematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorders; arthritis; inflammation.				
Homo sapiens.					
WO200164835-A2.					
26-FEB-2001; 2001WC-US004927.					
28-FEB-2000; 2000US-00515126.					
18-MAY-2000; 2000US-00577409.					
(HYSE-) HYSEQ INC.					
Tang YT, Liu C, Drmanac RT;					
WPI: 2001-514838/56;					
N-PADB; AAI62118.					
Isolated nucleic acids and polypeptides, useful for preventing diagnosing and treating e.g. leukemia, inflammation and immune disorders.					
Claim 20; SEQ ID NO 16079; 1399pp + Sequence listing; English.					
The invention relates to human polynucleotides (AAI79941-AA193841) and the encoded proteins (AAO0010-AA013910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences					

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: November 16, 2004, 12:58:46 ; Search time 167 Seconds

(without alignments)  
4272.243 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 1093  
Sequence: 1 gttcaaaatacatttca.....aaaaaaaaaaaaaaaa 620

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 3650362

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10822496/runat\_16112004\_100129\_24158/app\_query.fasta-1.775  
-DB=uniprot\_02 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=35 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10822496 @CCN 1 1 244 @runat\_16112004\_100129\_24158 -NCPU=6 -ICU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOG  
-DEVTIMEOUT=120 -WAKN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Uniprot\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No:	Score	Query Match length	ID	Description
C 1	95	8.3	725 2	Q9CV93 mus musculus
C 2	94.5	8.6	315 2	Q6IGT5 drosophila
C 3	94	8.2	4782 2	Q8K1G6 mus musculus
C 4	91.5	8.0	256 2	Q96177 plasmidum
C 5	91	8.0	5703 1	MUS8 HUMAN
C 6	89.5	7.8	669 2	Q8N4X0 mus musculus
C 7	89.5	7.7	1070 2	Q96JG5 mus musculus
C 8	88.5	7.7	3703 1	ABF1 HUMAN
C 9	86.5	7.6	446 2	Q95ND3 mus musculus
C 10	86.5	7.6	609 2	Q959C3 mus musculus
C 11	86.5	7.6	684 2	Q7F7M1 cyrtax sativ
C 12	86.5	7.6	684 2	Q9M577 cyrtax sativ
C 13	86	7.5	700 2	Q8NFK3 cyrtax sativ
C 14	84.5	7.4	579 2	Q7QH05 cyrtax sativ
C 15	84.5	7.7	782 1	L100 ADE12
C 16	84.5	7.7	782 2	Q7M6E4 human adeno

## ALIGNMENTS

RESULT 1	ID	Q9CV93	PRELIMINARY;	PRT;	725 AA.
AC	Q9CV93	01-JUN-2001 (TREMBLREL. 17, Created)			
DT	01-JUN-2001 (TREMBLREL. 17, Last sequence update)				
DT	01-MAR-2004 (TREMBLREL. 26, Last annotation update)				
DE	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:230002104 product:stimilar to MUCIN (Fragment).				
GN	Name=Muc5b;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Tongue;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	RIKEN FANTOM Consortium;				
RT	"Functional annotation of a full-length mouse cDNA collection.";				
RL	Nature 409:685-690(2001).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Tongue;				
RA	The FANTOM Consortium;				
RT	"Analysis of the mouse transcriptome based on functional annotation of				
RL	60,770 full-length cDNAs.";				
RN	Nature 420:563-573(2002).				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Tongue;				
RX	MEDLINE=20499374; PubMed=11042159;				

17	84.5	7.7	990 2	Q6NR11	Q6NR11 drosophila
18	84.5	7.7	990 2	AAQ23591	AAQ23591 drosophila
19	84.5	7.7	1052 2	Q9VF12	Q9VF12 drosophila
20	84.5	7.7	1138 2	Q9VF13	Q9VF13 drosophila
21	84	7.3	381 2	Q915P2	Q915P2 pseudomonas
22	84	7.3	2848 2	Q6W705	Q6W705 cyrtax sativ
23	84	7.3	2848 2	AAQ76546	AAQ76546 cyrtax sativ
24	83.5	7.3	404 2	Q6Z0K7	Q6Z0K7 cyrtax sativ
25	83.5	7.3	404 2	BAC87616	BAC87616 cyrtax sativ
26	83.5	7.3	611 2	Q959A5	Q959A5 galaxias an
27	83.5	7.6	824 2	Q6C200	Q6C200 cyrtax sativ
28	83	7.3	1707 2	Q6P9K2	Q6P9K2 mus musculus
29	83	7.3	2192 2	AAH60729	AAH60729 mus muscu
30	83	7.3	3726 1	ABF1 MOUSE	ABF1 MOUSE
31	83	7.3	233 2	Q9UFP2	Q9UFP2 mus musculus
32	82.5	7.5	289 2	Q9NUN6	Q9NUN6 mus musculus
33	82.5	7.5	385 2	Q8HVC5	Q8HVC5 mus musculus
34	82.5	7.5	533 2	Q9H8Z2	Q9H8Z2 mus musculus
35	82.5	7.5	584 2	Q6INS4	Q6INS4 mus musculus
36	82.5	7.5	584 2	AAH72453	AAH72453 mus sapi
37	82.5	7.2	611 2	Q959B3	Q959B3 galaxias br
38	82.5	7.2	1740 2	Q6ENK5	Q6ENK5 cyrtax sativ
39	82.5	7.5	325 1	HMGL_RAT	HMGL_RAT
40	82	7.2	325 2	AAH61797	AAH61797 rattus norv
41	82	7.2	358 2	Q6N172	Q6N172 corynebacte
42	82	7.2	358 2	CAE49422	CAE49422 corynebacte
43	82	7.2	358 2	P93520	P93520 zea mays (m
44	82	7.2	625 2		
45	82	7.5	625 2		

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630 (2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Tongue;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe J.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771 (2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Tongue;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hasegawa T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imosani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Kono H., Konda M., Koya S., Kurahara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,  
 RA Tejima Y., Toyota T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK009011; BAB26024.1; -  
 DR MGD: MGI:1921430; Muc5p.  
 DR GO: GO:0005615; C:extracellular space; IEA.  
 DR GO: GO:0005386; F:carrier activity; IEA.  
 DR GO: GO:000810; F:transport; IEA.  
 DR InterPro: IPR002919; Cysrich\_TITL.  
 DR InterPro: IPR006207; Cys\_Knot\_C.  
 DR InterPro: IPR009041; PMP\_SGCI.  
 DR InterPro: IPR000024; Serum\_albumin.  
 DR InterPro: IPR01007; VWF\_C.  
 DR InterPro: IPR01846; VWF\_D.  
 DR Pfam: PF00093; VWC; 1.  
 DR Pfam: PF00094; VMD; 1.  
 DR SMART: SM00041; CT; 1.  
 DR SMART: SM00214; VWC; 2.  
 DR PROSITE: PS00216; VMD; 1.  
 DR PROSITE: PS00212; ALBUMIN; 1.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS01208; VWF\_C1; UNKNOWN\_2.  
 DR PROSITE: PS0184; VWF\_C2; 2.  
 FT NON\_TER 1  
 SQ SEQUENCE 725 AA; 78872 MW; F2D96B2CAB7552B4 CRC64;  
 Alignment Scores:  
 Pred. No.: 2.2 Length: 725  
 Score: 95.00 Matches: 62  
 Percent Similarity: 29.66% Conservative: 24  
 Best Local Similarity: 21.38% Mismatches: 67  
 Query Match: 8.30% Indels: 137  
 Db: 2 Gaps: 16  
 US-10-822-496-5 (1-620) x G9CV93 (1-725)  
 QY 574 AGTTCCTTTATTCCTGCGCTCCCGCTTATCTGCGCTTAT----- 529  
 Db 229 SerSerCysLeu-AlaGlnProSerPro-----LeuProSerThrThrProGlnThrPr 246  
 QY 528 -GTGCGCTTCTCTCTACTTCCCGCGCTCCCTCCATATCAGCTGCCACATTTGTTCT 470

Db 246 oValSerSerThrLeuThrThrProCysProSerAlaProLeuCysGluLeuMetLe 266  
 QY 469 TTCTGCTCTCATTCACAAAGCGATTAAGGATGCGCT----- 433  
 Db 266 uSerGlnValPheAlaGluCysHisArgLeuLeuProProAspAlaPhePheArgSerCy 286  
 QY 432 -----GCTCTATCCACCAACTGCTGGATA 407  
 Db 286 vAlaSerAspHisCysAsnAlaAsnIleThrAspMetLeuCysGlnSerLeuGlnAlaTy 306  
 QY 406 TCGG-----TCCAACTTGAAGTCACATGATTCACAAAGCA----- 367  
 Db 306 zAlaAlaLeuCysArgAlaGlnGlyValCysThrAsnThrArgAsnAlaThrGlyGlyLe 326  
 QY 366 -----TCGGCTCCGCTCTC-----ATCCAC----- 346  
 Db 326 uCysAspLeuSerCysProProThrIlyGluTyArgProCysGlyProLeuHisProAl 346  
 QY 346 ----- 346  
 Db 346 aSerCysAsnSerArgThrGlnAspSerThrThrGlyMetLeuAlaGluGlyCysPheCy 366  
 QY 346 ----- 346  
 Db 366 sProGlnAsnGlnLeuLeuPheAsnSerArgMetAspIleCysValSerGluCysProCy 386  
 QY 345 -----GTAGAGAGAGCTCTCCAAATAG-----GACTGG-----TGCTACAC 308  
 Db 386 sValGlyProAspGlyLeuProLysPheProGlyGlnHisThrIleSerAsnGlyGlnAl 406  
 QY 307 GTGTGTCAAGGGGTACCGCGCTGGCGCATGTTGAC-----GCTCGAGCTGGG 263  
 Db 406 aCysValCys-----GluProGlySerValSerValGlnCysGluProValLysCysG1 424  
 QY 262 TGAAGGTCACAAAGCAATGATTCGGATTTGCTGATGCTGCTGCTGCTGCTGCTGCTG 203  
 Db 424 uSerGlnAspLysProProGlnCysThrGlnAlaGlyPheValThrValThrThrPro-- 443  
 QY 202 CAGTTGGAAGATATATTCATACATGATGAAACATTCGACCTATTAAACATTCCTGAG 143  
 Db 444 -----SerAlaAspAsn----- 447  
 QY 142 AGAAGAGTGTCTGTTCTCCCGCGCTA-----TCCAGCACTTCCCAAA----- 100  
 Db 448 -----ProCysCysProGlnThrLeuCysValCysAsnSerThrThrCysProLysSerLe 466  
 QY 99 -----TATCTAGAGAGAGCTTGTCTATA 74  
 Db 466 uProThrCysGluProGlyTyrgIleuValGlnThrHisGluAsnSerSerCysCysBr 486  
 QY 73 CAGTTATAC---AGTCTAGATATGT 49  
 Db 486 oSerTySerCysArgProArgLeuCys 495  
 RESULT 2  
 Q6IGTS PRELIMINARY; PRT; 315 AA.  
 AC 06IGTS;  
 DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
 DE HD005285.  
 OS ORFNames=HD005285.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 NC NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX PubMed=14709175;  
 RA Hild M., Beckmann B., Haas S., Koch B., Solovyev V., Busold C.,

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM nucleic - protein search, using frame\_plus.n2p model

Run on: November 16, 2004, 13:06:41 ; Search time 36 Seconds  
(without alignments)  
3314.133 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 1093

Sequence: 1 gttcaataataacattttaa.....aaaaaaaaaaaaaaaaaaaaa 620

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 6.0 , Xgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame\_plus.n2p.model -DEV=slp  
-Q=/cgn2/USFTO\_spool\_p/US10822496/runat.16112004.100130.24178/app.query.fasta.1.775  
-DB=PIR\_79 -QMT=faetan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFN=pcio -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US10822496 @CGN 1.1 63 @runat.16112004.100130.24178 -NCPU=6 -ICPU=3  
-NOR MMP= -LARGEJOB= -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAPN TIMEOUT=30 -THRAPDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELDP=6 -DELEXT=7

## Database :

PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	88.5	7.7	2783	1 A41948	alpha-fetoprotein
C 2	84.5	7.7	782	2 S33945	late protein, 100K
C 3	84	7.3	381	2 A83561	probable type II s
C 4	82.5	7.3	233	2 T17218	hypothetical prote
C 5	82	7.5	625	2 T02033	calcium/calmodulin
C 6	81.5	7.1	391	1 S15617	E2 protein - human
C 7	81.5	7.1	496	2 TS2132	probable sugar tra
C 8	81	7.1	427	2 S60742	alpha-transinducin
C 9	81	7.1	427	2 UC2203	U48h protein - Ma
C 10	81	7.1	3014	1 JG5620	genome polyprotein
C 11	80.5	7.0	587	2 H72647	hypothetical prote
C 12	79.5	6.9	587	2 T16867	probable cytochrom
C 13	79.5	6.9	354	2 T22274	hypothetical prote
C 14	79.5	7.3	1199	2 T15826	hypothetical prote

C 15	79	7.2	385	2 H89046	protein C10G8.8 (i
C 16	79	6.9	388	2 S36500	E2 protein - human
C 17	77	6.7	620	2 H69382	ABC transporter, A
C 18	76.5	6.7	655	2 T16538	hypothetical prote
C 19	76.5	6.7	730	2 A96768	protein T4012.13 (
C 20	76	6.6	321	2 P84597	probable prolone-r
C 21	76	6.6	329	2 B72390	m4C-methyltransfer
C 22	76	7.0	328	2 A72048	proteinkinase (impor
C 23	76	6.6	1172	1 TSHD2	chromospondin 2 p
C 24	75.5	6.6	347	2 A75483	hypothetical prote
C 25	75.5	6.9	348	2 S75601	ribonuclease H (EC
C 26	75.5	6.6	402	2 S11773	polygalacturonase
C 27	75.5	6.6	1459	2 A12468	hypothetical prote
C 28	75	6.9	168	2 T40352	single-stranded DN
C 29	75	6.6	1056	2 A53767	mucln MUC5B, trach
C 30	75	6.6	1194	2 B36624	hypothetical prote
C 31	74.5	6.5	213	2 F84532	hypothetical prote
C 32	74.5	6.5	378	2 T46268	brevican precursor
C 33	74.5	6.5	515	2 S53525	G protein-coupled
C 34	74.5	6.5	1105	2 T18295	Ap-3 adaptor compl
C 35	74.5	6.8	2288	2 T29999	hypothetical prote
C 36	74.5	6.8	2477	2 S14428	fibronectin precu
C 37	74	6.8	298	2 JC7870	ribonuclease Ie3 -
C 38	74	6.5	325	2 I57009	3-Hydroxy-3-methyl
C 39	74	6.5	340	2 E22360	Ig alpha-2 chain C
C 40	74	6.5	750	2 T42614	probable envelope
C 41	74	6.5	1246	2 C36806	hypothetical prote
C 42	74	6.5	1265	2 S17675	valine-tRNA ligase
C 43	74	6.5	1496	2 T40016	conserved hypothet
C 44	74	6.5	2325	2 A61208	chondroitin sulfat
C 45	73.5	6.4	262	2 T14610	hypothetical prote

## ALIGNMENTS

RESULT 1  
A41948  
alpha-fetoprotein enhancer-binding protein - human  
N:Alternate names: A7BFL protein  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #ext\_change 15-Oct-1999  
C:Accession: A41948  
R:Morinaga, T.; Yasuda, H.; Hashimoto, T.; Higashio, K.; Tamaoki, T.  
Mol. Cell. Biol. 11, 6041-6049, 1991  
A:Title: A human alpha-fetoprotein enhancer-binding protein, A7BFL, contains four homeod  
A:Reference number: A41948; MUID:92049333; PMID:1719379  
A:Accession: A41948  
A:Molecule type: mRNA  
A:Residues: 1-2783 <MOR>  
A:Cross-references: GB:D10250; GB:D90395; NID:9219429; PIDD:BA01095.1; PTD:9219430  
A:Note: sequence extracted from NCBI backbone (NCBI:66271, NCBI:66276)  
C:Genetics:  
A:Gene: GDB:A7BFL  
A:Cross-references: GDB:392090; OMIM:104155  
A:Map position: 16q22.3-16q23.1  
C:Superfamily: alpha-fetoprotein enhancer-binding protein; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger  
P:2,94/Region: zinc finger CCH motif  
F:128-150/Region: zinc finger CCH motif  
F:176-198/Region: zinc finger CCH motif  
F:311-332/Region: zinc finger CCH motif  
F:340-361/Region: zinc finger CCH motif  
F:448-471/Region: zinc finger CCH motif  
F:489-509/Region: zinc finger CCH motif  
F:517-538/Region: zinc finger CCH motif  
F:633-655/Region: zinc finger CCH motif  
F:684-706/Region: zinc finger CCH motif  
F:719-773/Region: serine/threonine-rich  
F:809-958/Region: glutamine-rich  
F:1071-1092/Region: zinc finger CCH motif  
F:1117-1211/Region: proline-rich  
F:1232-1288/Domain: homeobox homology <HOX1>  
F:1329-1385/Domain: homeobox homology <HOX2>

100

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 16, 2004, 13:19:42 : Search time 122 seconds  
(without alignments)  
3596.195 Million cell updates/sec

Title: US-10-822-496-5  
Perfect score: 1093  
Sequence: 1 gttcaataatcattttaa.....aaaaaaaaaaaaaa 620

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 156869 segs, 353819137 residues

Total number of hits satisfying chosen parameters: 3137398

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame-n2p-model -DEV=x1p  
-Q/cgnt1/usfro.spool.p/US10822496/runatc.16112004.100132.24268/app.query.fasta.1.775  
-DB=Published.Applications\_AA -QMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=10sum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=ext -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -MSR=US10822496 @CGN.1.1.199 @runatc.16112004.100132.24268  
-NCU=6 -ICPU=3 -NO USER -IAREQUERY -NEG SCORES=0 -WAIT -DSELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database : Published.Applications\_AA:

1: /cgnt2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgnt2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgnt2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgnt2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgnt2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgnt2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgnt2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgnt2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgnt2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgnt2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgnt2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
12: /cgnt2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgnt2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgnt2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgnt2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgnt2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgnt2\_6/ptodata/1/pubpaa/US10D\_NEW\_PUB.pep:\*  
18: /cgnt2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgnt2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
20: /cgnt2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
21: /cgnt2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
C 1	117.5	10.3	327	14	US-10-017-161-2258	Sequence 2258, Ap
C 2	117.5	10.3	327	14	US-10-292-798-1904	Sequence 1904, Ap
C 3	99	9.1	150	16	US-10-767-701-45440	Sequence 45440, A
C 4	92	8.0	153	16	US-10-437-963-156345	Sequence 156345,
C 5	91.5	8.0	137	15	US-10-424-599-145989	Sequence 145989,
C 6	89	7.8	159	17	US-10-425-115-23612	Sequence 23612,
C 7	88.5	8.1	236	16	US-10-437-963-149326	Sequence 149326,
C 8	88.5	8.1	401	16	US-10-767-701-44461	Sequence 44461, A
C 9	88.5	7.7	2783	9	US-09-816-6699-14	Sequence 14, Appl
C 10	88.5	7.7	2783	16	US-10-408-765A-2228	Sequence 2228, Ap
C 11	88	7.7	124	16	US-10-437-963-177169	Sequence 177169,
C 12	87.5	7.6	128	17	US-10-425-115-241376	Sequence 241376,
C 13	87.5	7.6	189	17	US-10-425-115-231208	Sequence 231208,
C 14	86.5	7.6	102	17	US-10-425-115-230642	Sequence 230642,
C 15	86.5	7.6	901	16	US-10-437-963-170537	Sequence 170537,
C 16	86	7.5	106	15	US-10-424-599-22677	Sequence 22677,
C 17	86	7.5	657	15	US-10-294-006-13	Sequence 13, Appl
C 18	86	7.5	895	15	US-10-294-006-35	Sequence 35, Appl
C 19	85	7.4	98	15	US-10-424-599-260338	Sequence 260338,
C 20	85	7.4	203	15	US-10-424-599-260338	Sequence 260338,
C 21	84.5	7.4	132	17	US-10-425-115-219535	Sequence 219535,
C 22	84.5	7.4	286	17	US-10-425-115-230494	Sequence 230494,
C 23	84	7.3	111	16	US-10-437-963-204029	Sequence 204029,
C 24	84	7.3	160	16	US-10-437-963-204029	Sequence 204029,
C 25	83.5	7.3	246	14	US-10-017-161-2056	Sequence 188033,
C 26	83.5	7.3	246	14	US-10-292-798-1702	Sequence 1702, Ap
C 27	83.5	7.3	274	14	US-10-128-714-3124	Sequence 3124, Ap
C 28	83.5	7.3	274	14	US-10-128-714-3124	Sequence 3124, Ap
C 29	83.5	7.3	341	14	US-10-091-438-166	Sequence 166, App
C 30	83.5	7.3	1057	14	US-10-189-971-6	Sequence 6, Appl
C 31	83.5	7.3	1192	14	US-10-189-971-18	Sequence 18, Appl
C 32	83.5	7.3	1251	14	US-10-189-971-20	Sequence 20, Appl
C 33	83.5	7.3	1432	14	US-10-189-971-24	Sequence 24, Appl
C 34	83.5	7.3	1512	14	US-10-189-971-8	Sequence 8, Appl
C 35	83.5	7.3	1512	14	US-10-189-971-10	Sequence 10, Appl
C 36	83.5	7.3	1512	14	US-10-189-971-14	Sequence 14, Appl
C 37	83.5	7.3	1570	14	US-10-189-971-12	Sequence 12, Appl
C 38	83.5	7.3	1570	14	US-10-189-971-12	Sequence 12, Appl
C 39	83.5	7.3	1593	14	US-10-189-971-4	Sequence 4, Appl
C 40	83.5	7.3	1628	14	US-10-189-971-2	Sequence 2, Appl
C 41	83	7.6	253	15	US-10-424-599-157200	Sequence 157200,
C 42	82.5	7.5	77	15	US-10-424-599-235133	Sequence 235133,
C 43	82.5	7.2	143	17	US-10-425-115-344539	Sequence 344539,
C 44	82.5	7.5	533	15	US-10-051-874-137	Sequence 137, App
C 45	82.5	7.5	533	16	US-10-408-765A-1052	Sequence 1052, Ap

## ALIGNMENTS

RESULT 1  
US-10-017-161-2258  
; Sequence 2258, Application US/10017161  
; Publication No. US20030143668A1  
; GENERAL INFORMATION:  
; APPLICANT: SUMA, YAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABEURANT, HIROYUKI  
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 084335/0152  
; CURRENT APPLICATION NUMBER: US/10/017,161  
; PRIOR APPLICATION NUMBER: 2002-12-18  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2430  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2258  
; LENGTH: 327  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: November 16, 2004, 13:07:01 ; Search time 29 Seconds

28335.666 Million cell updates/sec

```
Title: US-10-822-496-5
Perfect score: 1093
Sequence: 1 gtccaaataacattctaa.....aaaaaaaaaaaaaaaaaa 620
```

Scoring table:		BLOSUM62
Xgapop	10.0	Xgapext 0.5
Ygapop	10.0	Ygapext 0.5
Fgapop	6.0	Fgapext 7.0
Delop	6.0	Delext 7.0

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 956278

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match 0%
	Maximum Match 100%

Listing first 45 summaries

### Command line parameters

```

-O=csg2.1/USPO.spool/p/US10822496/runat.16112004.100130.24191/app.qerry.faste.1.775
-DB=csgr2.Patents_AA_OPMT=fastean.SUPPLY=adp.ra1.MATMATHC=0.1.-LOOPY=0
-LOOPEXT=0.-UNITS=bits.-START=1.-END=-1.-MATRIX=blotsun62.-TRANS=human40.cdi
-LIST=45.-DCCALL=LOCAL=100.-THR SCORE=pct.-THR MAX=100.-THR MIN=0.-ALIGN=15
-MODE=LOCAL.-OUTPMF=pcio.-NOS=ext.-HEAPSIZE=500.-MINLEN=0.-MAXLEN=20000000000
-USER=US10822496.csgn.1.1.46.@runat.16112004.100130.24191.-NCPU=6.-ICPU=3
-NO MMAP.-LAAGBDEXY.-NEG SCORES=0.-WANT DSBLOCK=100.-LONGLOG
-DEV TIMEOUT=120.-WARN TIMEOUT=10.-KAPAD=10.-XGAPEXT=0.5.-FGAPAD=6
FGAPEXT=7.-XGAPO=10.-YGAPEXT=0.5.-IDEP=6.-IDEXMT=7

```

Database :

```

1: /cgn2_6/ptodata1/1/aa/5A COMB pep: *
2: /cgn2_6/ptodata1/1/aa/5B COMB pep: *
3: /cgn2_6/ptodata1/1/aa/6A COMB pep: *
4: /cgn2_6/ptodata1/1/aa/6B COMB pep: *
5: /cgn2_6/ptodata1/1/aa/PCTUS COMB pep: *
6: /cgn2_6/ptodata1/1/aa/backfilest1.pep: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	90.5	8.3	112	4	US-09-543-6814-8319		Sequence 8319, Apr
2	87	8.0	171	4	US-09-469-0394-8235		Sequence 8235, Apr
3	85.5	7.5	515	4	US-08-930-9564-12		Sequence 12, Apr
4	85	7.4	169	4	US-09-252-9914-1686		Sequence 1686, Apr
5	82.5	7.5	178	4	US-09-489-0394-8235		Sequence 8235, Apr
6	82.5	7.2	258	4	US-09-252-9914-16825		Sequence 16825, Apr
7	82	7.5	625	3	US-09-347-801-18		Sequence 18, Apr
8	82	7.5	625	4	US-09-854-731-18		Sequence 18, Apr
9	81.5	7.1	142	4	US-09-248-7964-2139		Sequence 2139, Apr
10	80.5	7.4	722	4	US-09-252-9914-22811		Sequence 22811, Apr
11	80	7.3	160	4	US-09-543-6814-8310		Sequence 8310, Apr
12	80	7.3	182	4	US-09-489-0394-8301		Sequence 8301, Apr

C 13	77.5	7.1	65	4	US-09-621-976-5159	Sequence 5159, App
C 14	77.5	7.1	803	4	US-09-154-750A-85	Sequence 85, App
C 15	77.5	7.1	803	4	US-09-665-479A-12	Sequence 12, App
C 16	77	6.7	236	4	US-09-270-767-60936	Sequence 60936, A
C 17	77	6.7	460	4	US-09-270-767-33708	Sequence 33708, A
C 18	77	6.7	460	4	US-09-270-767-48925	Sequence 48925, A
C 19	77	6.7	520	4	US-09-270-767-45430	Sequence 45430, A
C 20	76	7.0	111	4	US-09-543-681A-8332	Sequence 8332, App
C 21	76	7.0	418	4	US-09-252-991A-23035	Sequence 23035, A
C 22	76	7.0	357	4	US-09-388-413B-6	Sequence 6, App11
C 23	76	7.0	357	4	US-10-233-745-6	Sequence 6, App11
C 24	76	6.6	1172	1	US-08-313-288B-19	Sequence 19, App11
C 25	75.5	6.6	346	4	US-10-140-002-432	Sequence 432, App11
C 26	75.5	6.9	348	3	US-09-203-0716-3	Sequence 3, App11
C 27	75.5	6.9	348	3	US-09-684-254-3	Sequence 3, App11
C 28	75.5	5.9	348	4	US-09-409-926-3	Sequence 3, App11
C 29	75	5.6	230	4	US-09-270-767-43372	Sequence 43372, A
C 30	74.5	6.8	134	4	US-09-252-991A-27332	Sequence 27332, A
C 31	74.5	6.5	227	2	US-08-580-545B-6	Sequence 6, App11
C 32	74.5	6.5	227	3	US-09-262-653A-6	Sequence 6, App11
C 33	74.5	6.5	378	1	US-08-225-477B-9	Sequence 9, App11
C 34	74.5	6.5	378	1	PCT-US95-04353-9	Sequence 9, App11
C 35	74.5	6.8	458	4	US-09-252-991A-19911	Sequence 19911, A
C 36	74.5	6.5	515	4	US-09-170-496D-104	Sequence 104, AppA
C 37	74.5	6.5	515	4	US-09-170-496D-220	Sequence 220, App11
C 38	74.5	6.5	567	4	US-09-514-245-6	Sequence 6, App11
C 39	74	6.5	178	4	US-09-673-395A-6	Sequence 330, App11
C 40	74	6.8	335	4	US-09-252-991A-23038	Sequence 23038, A
C 41	74	6.5	483	3	US-09-045-672A-5	Sequence 5, App11
C 42	74	6.5	487	4	US-09-800-729-145	Sequence 145, App1
C 43	74	6.5	583	4	US-09-514-245-16	Sequence 16, App1
C 44	74	6.5	1157	4	US-09-538-052-1328	Sequence 1328, App
C 45	74	6.5	1220	1	US-08-156-232-43	Sequence 43, App

## ALIGNMENTS

RESULT 1  
US-09-543-681A-8319  
; Sequence 8319, Application US/09543681A

APPLICANT: GARY BRETON

! TITLE OF INVENTION:   DIAGNOSTICS AND THERAPEUTICS

CURRENT APPLICATION NUMBER: US/09/543,681A

PRIOR APPLICATION NUMBER: US 60/128,706

NUMBER OF SEQ ID NOS:  
SEQ ID NO 8319

TYPE: PRT  
LENGTH: 112

US-09-543-681A-8319

Alignment Scores:

Score:	90.50	Matches:	23
--------	-------	----------	----

Best Local Similarity:	40.35%	Mismatches:	22
Overall Match:	9.28%	Indels:	5

DB: 4  
Cups:

US-10-822-496-5 (1-620) X US-09-543-681A-8319 (1-112)

[illegible]

*[Illegible vertical text]*

[illegible]

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/930,996A
  FILING DATE: 09-DEC-1997
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/US96/05272
    FILING DATE: 15-APR-1996
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: IL 113,373
      FILING DATE: 13-APR-1995
  INFORMATION FOR SEQ ID NO: 12:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 515 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
  MOLECULE TYPE: protein
  US-08-930-996A-12

Alignment Scores:
  Pred. No.: 0.359
  Score: 85.50
  Percent Similarity: 50.65%
  Best Local Similarity: 32.47%
  Query Match: 7.47%
  DB: 3
  Gaps: 3

US-10-822-496-5 (1-620) x US-08-930-996A-12 (1-515)
QY 583 TCCACATGCAAGTTCCTGTTTATTCCTTACCCCTTATCTGACCTATGTTGC 524
Db 413 SerHisCysProAsn-LeuGlnSerLeuProGluSerAlaLeuLeuSerSerLeuSerHi 432
QY 523 CCTTCTTCACATTCGCCGCCCTGCCT-----CC 494
Db 432 sLeuAspIleSerHis-----CysProAsnLeuGlnSerLeuAlaGluSerAlaLeuPr 450
QY 493 AATACAGCTGTGCCCATGTTCTTTCT-----GGTTCATTCGAAAGCCGATAA 443
Db 450 oSerSerLeuSerHisLeuThrIleSerHisCysProAsnLeuHisLeuSerLeuGlu 470
QY 442 GGGAGTCCCTGCTCTTAATCCAGCACTGCTCGGATATGCGTGCACCTT 394
Db 470 scGlyAcProSerSerLeuSerIleSerIleSerIleSerIleSerIleSerIle 486

RESULT 4
US-09-252-991A-16868
  Sequence 16868, Application US/0925291A
  Patent No. 6551795
  GENERAL INFORMATION:
    APPLICANT: Marc J. Rubenfield et al.
    TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
    FILE REFERENCE: 107156.136
    CURRENT APPLICATION NUMBER: US/09/252,991A
    CURRENT FILING DATE: 1999-02-18
    PRIOR APPLICATION NUMBER: US 60/074,788
    PRIOR FILING DATE: 1998-02-18
    PRIOR APPLICATION NUMBER: US 60/094,190
    PRIOR FILING DATE: 1998-07-27
    NUMBER OF SEQ ID NOS: 33142
    SEQ ID NO 16868
    LENGTH: 169
    TYPE: PRT
  ORGANISM: Pseudomonas aeruginosa
  US-09-252-991A-16868

Alignment Scores:
  Pred. No.: 0.249
  Score: 85.00
  Percent Similarity: 33.85%
  Best Local Similarity: 27.60%
  Query Match: 7.43%
  Length: 169
  Matches: 53
  Conservative: 12
  Mismatches: 58
  Indels: 69

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 16, 2004, 12:58:01, Search time 125 Seconds

(without alignments)  
3558.595 Million cell updates/sec

Title: US-10-822-496-5  
Perfect score: 1093  
Sequence: 1 gtccaaataatacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table:

BLASTN62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Xgapop 6.0, Ygapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2002273 seqs, 35872929 residues

Total number of hits satisfying chosen parameters: 4004546

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_n2p\_model -DEV=ralp  
-Q=/cgrt\_1/USPTO.spool\_y/US10822496/funat\_16112004\_100129\_24151/app\_query.fasta\_1.775  
-DB=Geneseq\_23Sep04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10822496 @CGN 1.1 224 @runat 16112004 100129 24151 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUTERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -RAPOP=6  
-FgapExt=7 -YgapExt=10 -YgapExt=0.5 -DELOP=6 -DELEX=7

Database: A\_Geneseq\_23Sep04:\*

1: Geneseqp1980s:\*\n2: Geneseqp1990s:\*\n3: Geneseqp2000s:\*\n4: Geneseqp2001s:\*\n5: Geneseqp2002s:\*\n6: Geneseqp2003as:\*\n7: Geneseqp2003bs:\*\n8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	117.5	10.3	327	7	ADc87451 Human GPC
C 2	91	8.0	4315	5	ABP43908 MUC5B par
C 3	91	8.0	5703	9	AD123265 Human MUC
C 4	90.5	8.3	112	7	ADf08034 Bacterial
C 5	89.5	7.8	1030	6	ABP58606 Human can
C 6	89.5	7.8	1070	7	ADP55479 Human nov
C 7	88.5	7.7	2763	2	AAR23963 Human pol
C 8	88.5	7.7	2763	2	AAR23962 AFP-1 (AI
C 9	88.5	7.7	2763	4	AAB82946 Human and
C 10	88.5	7.7	2769	7	ADJ70422 Human hea

11	87	8.0	171	7	ABO61718 Klebsiell
C 12	86	7.5	657	6	ABG70519 Human pol
C 13	86	7.5	657	8	ADJ92551 Human pro
C 14	86	7.5	895	8	ADJ92573 Human pro
C 15	85.5	7.5	113	4	ABG09348 Novel hum
C 16	85.5	7.5	719	4	ABG04957 Novel hum
C 17	85	7.4	169	7	ABO68122 Pseudomon
C 18	85	7.4	203	4	ABE11632 Human dop
C 19	84.5	7.4	145	4	ABG28355 Novel hum
C 20	84.5	7.4	1058	4	ABE71401 Drosophil
C 21	83.5	7.3	246	7	ADc87249 Human GPC
C 22	83.5	7.3	274	6	ABJ26066 Apepergill
C 23	83.5	7.3	274	6	ABJ25466 Apepergill
C 24	83.5	7.6	341	4	AAU18021 Human imm
C 25	83.5	7.6	341	7	ADB31645 Human nov
C 26	83.5	7.3	1057	6	ABP97370 Human kie
C 27	83.5	7.3	1192	6	ABP97376 Human kie
C 28	83.5	7.3	1207	6	ABP97377 Human kie
C 29	83.5	7.3	1251	6	ABP97375 Human kie
C 30	83.5	7.3	1342	6	ABP97379 Human kie
C 31	83.5	7.3	1477	6	ABP97371 Human kie
C 32	83.5	7.3	1512	6	ABP97372 Human kie
C 33	83.5	7.3	1515	6	ABP97374 Human kie
C 34	83.5	7.3	1570	6	ABP97373 Human kie
C 35	83.5	7.3	1593	6	ABP97369 Human kie
C 36	83.5	7.3	1628	6	ABP97368 Human kie
C 37	83	7.3	153	4	AAU31783 Novel hum
C 38	82.5	7.2	160	4	ABG27888 Novel hum
C 39	82.5	7.2	178	7	ABO61719 Klebsiell
C 40	82.5	7.2	258	7	ABO71079 Pseudomon
C 41	82.5	7.2	289	4	AAAB93619 Human pro
C 42	82.5	7.2	291	6	ABR01685 Human pol
C 43	82.5	7.5	519	4	AAAD0360 Human cel
C 44	82.5	7.5	533	4	AAAB0463 Human cel
C 45	82.5	7.5	533	4	AAAB95321 Human pro

RESULT 1	ADc87451	standard: protein: 327 AA.	ALIGNMENTS
XX	ADc87451;		
XX	01-JAN-2004 (first entry)		
XX	Human GPCR protein SEQ ID NO:1904.		
XX	human; GPCR; guanosine triphosphate-binding protein coupled receptor;		
XX	gene therapy.		
OS	Homo sapiens.		
XX	EP1270724-A2.		
PN	02-JAN-2003.		
XX	16-JUN-2002; 2002EP-00013517.		
PF	18-JUN-2001; 2001JP-00246789.		
XX	(NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.		
PA	(ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.		
PI	Suwa M, Arai K, Akiyama Y, Aburatani H;		
XX	WPI: 2003-315783/31.		
DR	N-PSDB; ADc87450.		
XX	New polynucleotide, useful for preparing a composition for treating a		
PT	patient in need of increased or suppressed activity or expression of the		
PT	guanosine triphosphate-binding protein coupled receptor.		



Tue Nov 23 09:27:43 2004

us-10-822-496-5.01g.rst

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:42:52 ; Search time 2993 Seconds

(without alignments)  
7548.491 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620

Sequence: 1 gttcaaaatcacatttca.....aaaaaaaaaaaaaa 620

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 32822875 seqs, 1821985908 residues

Word size : 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database : EST:\*

1: gb\_esc1:.\*  
2: gb\_esc2:.\*  
3: gb\_esc3:.\*  
4: gb\_esc4:.\*  
5: gb\_esc5:.\*  
6: gb\_esc6:.\*  
7: gb\_esc7:.\*  
8: gb\_esc8:.\*  
9: gb\_esc9:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	4.7	543	7	CK360119 AGENCOURT
2	28	4.7	601	2	BF681816 602117624
3	28	4.5	199	1	A1365051 xrl2b04.x
4	28	4.5	285	2	BE049016 xre4f11.x
5	28	4.5	355	4	BM378616 MEST567-F
6	28	4.5	384	7	CN171083 AGENCOURT
7	28	4.5	406	4	BI510807 AGENCOURT
8	28	4.5	585	6	CD666951 NCST3b53
9	28	4.5	590	5	BM074026 MEST79-B1
10	28	4.5	621	4	CF515989 CAP002.1
11	28	4.5	676	7	CR695006 Terradon
12	28	4.5	910	3	CR673159 Terradon
13	28	4.5	921	3	CR673159 Terradon
14	28	4.5	949	7	CO757970 brain.EST
15	28	4.5	1101	9	AI063694 Drosophila
16	28	4.5	1104	9	CR679551 Terradon
17	28	4.4	99	7	CK736897 T95T2yko
18	27	4.4	165	4	BI813177 J005B08 O
19	27	4.4	171	1	A1105935 b03h08.P
20	27	4.4	186	6	CD722635 o12b06.y
21	27	4.4	211	7	N80246 za96b01.sl
22	27	4.4	212	6	BU531282 AGENCOURT
23	27	4.4	214	5	CD724006 c72b03.y
24	27	4.4	236	2	AM073961 xB05f06.x

25	27	4.4	236	4	BI512339	BI512339 BB160008B
26	27	4.4	250	7	CK678589	CK678589 2F101-P00
27	27	4.4	274	2	BF530231	BF530231 602071406
28	27	4.4	276	6	CD673937	CD673937 f805d07.x
29	27	4.4	292	6	CB709984	CB709984 AMGNNUC.U
30	27	4.4	293	6	CF202144	CF202144 KR890915N
31	27	4.4	308	6	CB048748	CB048748 NISC_g106
32	27	4.4	314	5	BU565918	BU565918 AGENCOURT
33	27	4.4	342	2	AM446955	AM446955 87884.MAR
34	27	4.4	349	1	AI022895	AI022895 0544f10.s
35	27	4.4	356	4	BG486009	BG486009 dab10b04.
36	27	4.4	357	7	CP972516	CP972516 EST1045.Z
37	27	4.4	380	5	BM197491	BM197491 BM192491
38	27	4.4	382	6	CB808280	CB808280 AMGNNUC.Y
39	27	4.4	383	1	AI185107	AI185107 GES1107.X
40	27	4.4	383	1	AI388033	AI388033 GH18764.X
41	27	4.4	384	4	BI979250	BI979250 fL85a08.Y
42	27	4.4	392	7	CK235610	CK235610 SB0100130
43	27	4.4	393	7	CN233362	CN233362 RJAI10C02
44	27	4.4	395	4	BM510577	BM510577 i47f09.x
45	27	4.4	398	4	BM511408	BM511408 i47f09.y

#### ALIGNMENTS

RESULT 1  
LOCUS CK360119 543 bp mRNA linear EST 23-DEC-2003  
DEFINITION AGENCOURT\_17116318 NIH\_ZGC\_4 Dario rerio cDNA clone IMAGE:7087989  
5', mRNA sequence.  
CK360119  
CK360119.1 GI:40326051  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Dario rerio (zebrafish)  
Dario rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 543)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: John Ngai, Univ of CA, Berkeley  
cDNA Library Preparation: Dr. Sumio Sugano  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LHAM14919 row: b column: 19  
High quality sequence, stop: 393.  
Location/Qualifiers  
1..543  
/organism="Dario rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:7087989"  
/lab\_host="DH10B TONR"  
/clone\_id="NIH\_ZGC\_4"  
/note="Organ: brain/CNS; Vector: pXEBIS-FU3; Site\_1:  
DraIII; Site\_2: DraIII"

#### ORIGIN

Query Match 4.7% Score 29; DB 7; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.0028;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 CGAAGGAGGAAAAAAAAAAAAAAAAAAAA 620  
 DB 282 CGAAGGAGGAAAAAAAAAAAAAAAAAAAA 310

RESULT 2  
 LOCUS Bf681816  
 DEFINITION 601 bp mRNA linear EST 21-DEC-2000  
 IMAGE:1468916 5', mRNA sequence.

ACCESSION Bf681816  
 VERSION Bf681816  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://mgi.nci.nih.gov/  
 1 (bases 1 to 601)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-rcmail.nih.gov  
 Tissue Procurement: M. Bento Soares, Ph.D.  
 CDNA Library Preparation: Bento Soares and M. Fatima Bonaldo  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.liml.gov  
 Place: LAM8476 row: p column: 05  
 High quality sequence start: 7  
 High quality sequence stop: 479.  
 Location/Qualifiers  
 1..601  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:13468916"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares mammary gland NMLMG"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN  
 Query Match 4.7%; Score 29; DB 2; Length 601;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 CGAAGGAGGAAAAAAAAAAAAAAAAAAAA 620  
 DB 424 CGAAGGAGGAAAAAAAAAAAAAAAAAAAA 452

RESULT 3  
 LOCUS A1365051  
 DEFINITION 199 bp mRNA linear EST 13-FEB-1999  
 IMAGE:1242404 5', mRNA sequence.

ACCESSION A1365051  
 VERSION A1365051  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 199)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-rcmail.nih.gov  
 Tissue Procurement: Christopher A. Morkaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at:  
 www.bio.liml.gov/bbtp/image/image.html  
 Insert length: 1431 Std Error: 0.00  
 Seq primer: -40bp from Glibco  
 High quality sequence stop: 189.  
 Location/Qualifiers  
 1..199  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1947343"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI-CGAP GC4"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN  
 Query Match 4.5%; Score 28; DB 1; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 0.01;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 593 GAACGAGGAAAAAAAAAAAAAAAAAAAA 620  
 DB 48 GAACGAGGAAAAAAAAAAAAAAAAAAAA 21

RESULT 4  
 LOCUS BE049016  
 DEFINITION 285 bp mRNA linear EST 08-JUN-2000  
 IMAGE:2764941 3', mRNA sequence.

ACCESSION BE049016  
 VERSION BE049016  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 285)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-rcmail.nih.gov  
 Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Krizman, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be



Tue Nov 23 09:27:42 2004

us-10-822-496-5.olg.rml

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:50:30 ; Search time 91 Seconds  
(without alignments)

4842.737 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620

Sequence: 1 gtcacaaataacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO\_NUC

Gapop 60.0, Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/1/ina/5A\_CONB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_CONB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_CONB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_CONB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCUTS\_CONB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles.seq:\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4.2	1384	3	US-09-227-357-142	Sequence 142, App
2	4.0	685	2	US-08-254-354-1	Sequence 1, Appl
3	4.0	685	2	US-08-254-354-3	Sequence 3, Appl
4	4.0	685	5	PCT-US95-06137-1	Sequence 1, Appl
5	4.0	685	5	PCT-US95-06137-3	Sequence 3, Appl
6	4.0	1361	4	US-09-614-912-37	Sequence 37, Appl
7	4.0	5207	4	US-09-858-664A-1	Sequence 1, Appl
8	4.0	5207	4	US-10-274-978-1	Sequence 1, Appl
9	4.0	5207	4	US-10-274-978-3	Sequence 3, Appl
10	3.9	53	4	US-09-621-976-11819	Sequence 11819, A
11	3.9	55	4	US-09-621-976-11651	Sequence 11651, A
12	3.9	57	4	US-09-513-998C-19821	Sequence 19821, A
13	3.9	59	4	US-09-621-976-11876	Sequence 11876, A
14	3.9	66	2	US-08-776-944-11	Sequence 11, Appl
15	3.9	69	4	US-09-621-976-11807	Sequence 11807, A
16	3.9	70	4	US-09-621-976-12454	Sequence 12454, A
17	3.9	74	4	US-09-621-976-12440	Sequence 12440, A
18	3.9	72	4	US-09-621-976-12390	Sequence 12390, A
19	3.9	77	4	US-09-621-976-12265	Sequence 12265, A
20	3.9	81	4	US-09-621-976-12198	Sequence 12198, A
21	3.9	81	4	US-09-621-976-12198	Sequence 12198, A
22	3.9	82	4	US-09-621-976-11689	Sequence 11689, A
23	3.9	82	4	US-09-621-976-11841	Sequence 11841, A
24	3.9	82	4	US-09-621-976-11864	Sequence 11864, A
25	3.9	82	4	US-09-621-976-11888	Sequence 11888, A
26	3.9	82	4	US-09-621-976-11944	Sequence 11944, A
27	3.9	82	4	US-09-621-976-11949	Sequence 11949, A

28	3.9	82	4	US-09-621-976-12015	Sequence 12015, A
29	3.9	82	4	US-09-621-976-12088	Sequence 12088, A
30	3.9	82	4	US-09-621-976-12103	Sequence 12103, A
31	3.9	82	4	US-09-621-976-12127	Sequence 12127, A
32	3.9	82	4	US-09-621-976-12131	Sequence 12131, A
33	3.9	82	4	US-09-621-976-12137	Sequence 12137, A
34	3.9	82	4	US-09-621-976-12322	Sequence 12322, A
35	3.9	82	4	US-09-621-976-12368	Sequence 12368, A
36	3.9	82	4	US-09-621-976-13499	Sequence 13499, A
37	3.9	82	4	US-09-621-976-13526	Sequence 13526, A
38	3.9	83	4	US-09-621-976-12087	Sequence 12087, A
39	3.9	83	4	US-09-621-976-12175	Sequence 12175, A
40	3.9	83	4	US-09-621-976-12195	Sequence 12195, A
41	3.9	83	4	US-09-621-976-12429	Sequence 12429, A
42	3.9	84	4	US-09-621-976-12450	Sequence 12450, A
43	3.9	84	4	US-09-621-976-11710	Sequence 11710, A
44	3.9	91	4	US-09-621-976-12161	Sequence 12161, A
45	3.9	98	4	US-09-621-976-11744	Sequence 11744, A

# ALIGNMENTS

RESULT 1  
US-09-227-357-142  
; Sequence 142, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; EARLIER APPLICATION NUMBER: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13664  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,916  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,930  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,918  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,920  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,733  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,795  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,919  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,928  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/055,722  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,723  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,948  
; EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-19  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 142  
LENGTH: 1384  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-227-357-142

Query Match 4.2% Score 26; DB 3; Length 1384;  
Best Local Similarity 100.0%; Pred. No. 0.0079;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 595 ACGAGAAAAAAAAAAAAAAAAAAAAA 620  
DB 1336 ACGAGAAAAAAAAAAAAAAAAAAAAA 1361

RESULT 2  
US-08-254-354-1  
Sequence 1, Application US/08254354  
Patent No. 5952210  
GENERAL INFORMATION:  
APPLICANT: Creely, David P.  
APPLICANT: Hauser, Scott D.  
APPLICANT: Welisch, Dean J.  
TITLE OF INVENTION: Nucleic Acids and Expression Vectors  
TITLE OF INVENTION: Encoding Human Leukotriene C4 Synthase, Antisense  
TITLE OF INVENTION: Oligonucleotides, and Methods of Use  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Scott B. Feder, G. D. Searle & Co., Corporate  
STREET: P. O. Box 5110  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60680  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/254,354  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Feder, Scott B.  
REGISTRATION NUMBER: 33,129  
REFERENCE/DOCKET NUMBER: 2800  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 470-6501  
TELEFAX: (708) 470-6881  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 685 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 87..539  
US-08-254-354-1

Query Match 4.0% Score 25; DB 2; Length 685;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 596 CGAGAAAAAAAAAAAAAAAAAAAAA 620  
DB 653 CGAGAAAAAAAAAAAAAAAAAAAAA 677

RESULT 3  
US-08-254-354-3/C  
Sequence 3, Application US/08254354  
Patent No. 5952210  
GENERAL INFORMATION:  
APPLICANT: Creely, David P.  
APPLICANT: Hauser, Scott D.  
APPLICANT: Welisch, Dean J.  
TITLE OF INVENTION: Nucleic Acids and Expression Vectors  
TITLE OF INVENTION: Encoding Human Leukotriene C4 Synthase, Antisense  
TITLE OF INVENTION: Oligonucleotides, and Methods of Use  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Scott B. Feder, G. D. Searle & Co., Corporate  
STREET: P. O. Box 5110  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60680  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/254,354  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Feder, Scott B.  
REGISTRATION NUMBER: 33,129  
REFERENCE/DOCKET NUMBER: 2800  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (708) 470-6501  
TELEFAX: (708) 470-6881  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 685 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ANTI-SENSE: YES  
US-08-254-354-3

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:42:52 ; Search time 422 Seconds

(without alignments)  
7712.420 Million cell updates/sec

Title: US-10-822-496-5  
Perfect score: 620  
Sequence: 1 gttcaaaaataacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001s:\*  
5: geneseqn2001ds:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003ds:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	4.4	103	ADCI7856	ADCI7856 Monobactam
2	27	4.4	165	ABZ73026	ABZ73026 Rice leaf
3	27	4.4	251	AAI13978	AAI13978 Human gen
4	27	4.4	473	ABV95307	ABV95307 Human pan
5	26	4.2	120	AAA89576	AAA89576 Exo20 nuc
6	26	4.2	157	ABP86076	ABP86076 Human cva
7	26	4.2	341	AAI95361	AAI95361 Cat flag
8	26	4.2	383	AAI19639	AAI19639 Human bre
9	26	4.2	792	AAI194836	AAI194836 Human neu
10	26	4.2	1054	AAH48740	AAH48740 Thraustoc
11	26	4.2	1384	AAI27442	AAI27442 Human sec
12	26	4.2	1384	ADA07321	ADA07321 Human DNA
13	26	4.2	1384	ADN41015	ADN41015 Novel hum
14	26	4.2	1462	ADN40910	ADN40910 Novel hum
15	26	4.2	1696	ABQ73684	ABQ73684 Human pot
16	26	4.2	2617	AAI170851	AAI170851 Human hep
17	26	4.2	11757	AAK85731	AAK85731 Human imm
18	25	4.0	325	ABV98209	ABV98209 Human pro
19	25	4.0	330	ACH31895	ACH31895 Human end
20	25	4.0	331	ACH36401	ACH36401 Human end
21	25	4.0	394	AAI10059	AAI10059 Human bre

22	25	4.0	467	ABV58112	ABV58112 Human pro
23	25	4.0	520	ABV56379	ABV56379 Human pro
24	25	4.0	587	AAH25771	AAH25771 Human cel
25	25	4.0	587	ABZ34920	ABZ34920 Human gen
26	25	4.0	587	ADP47748	ADP47748 CXB encod
27	25	4.0	603	AAH25770	AAH25770 Human cel
28	25	4.0	603	ADP47746	ADP47746 CK-HA enc
29	25	4.0	685	AAI11632	AAI11632 Leukotrin
30	25	4.0	685	AAI11631	AAI11631 Human leu
31	25	4.0	1027	AAI19818	AAI19818 Human pol
32	25	4.0	1034	AAI199531	AAI199531 Human pol
33	25	4.0	1361	ADN94237	ADN94237 Rice GA-2
34	25	4.0	1743	ABAO0091	ABAO0091 Breast/Co
35	25	4.0	1876	ACC43639	ACC43639 Nucleotid
36	25	4.0	2700	AAH33268	AAH33268 Human col
37	25	4.0	4175	AAI63285	AAI63285 CDNA encod
38	25	4.0	5207	AAI43909	AAI43909 Human kin
39	25	4.0	5207	AAI43908	AAI43908 Human kin
40	25	4.0	8106	ABX11641	ABX11641 Human ser
41	25	4.0	24120	ABX11642	ABX11642 Human ser
42	25	4.0	32134	AAI99172	AAI99172 Human exc
43	25	4.0	32134	AAI63522	AAI63522 Human kid
44	25	4.0	32192	AAI99173	AAI99173 Human exc
45	25	4.0	32192	AAI63523	AAI63523 Human kid

## ALIGNMENTS

RESULT 1	ADCI7856	ADCI7856 standard; RNA; 103 BP.
ID	ADCI7856	
AC	ADCI7856;	
DT	18-DEC-2003	(first entry)
DE	Monobactam related 81 biocatalyst subpopulation consensus SEQ ID NO:106.	
DS	ss; monobactam; antibacterial; PBPA; inhibitor;	
KM	methicillin resistant Staphylococcus aureus; MRSA; lactam antibiotic.	
OS	Synthetic.	
FX	Key	Location/Qualifiers
FT	modified_base	1..103
FT	FT	/*tag= a
XX	XX	/mod_base= 5-(4-pyridylmethyl)U
PN	WO2003051314-A2.	
PD	26-JUN-2003.	
XX	18-DEC-2002; 2002WO-US040739.	
PF	18-DEC-2001; 2001US-0340255P.	
PR	(INVE-) INVENTX INC.	
PA	Eaton B, Tarasow T, Nieuwlandt D, Dewey T;	
PI	WPI; 2003-618003/58.	
DR		
XX		
XX		
PT	New monobactam compounds used as antibacterial agents against e.g.	
PT	methicillin resistant Staphylococcus aureus.	
XX		
PS	Disclosure; SEQ ID NO 106; 64bp; English.	
XX		
CC	The invention relates to novel monobactam compounds. A compound of the	
CC	invention has antibacterial activity, and acts as a PBPA inhibitor. The	
CC	compounds are used as antibacterial agents. The monobactam compounds	
CC	restore sensitivity of methicillin resistant Staphylococcus aureus to	
CC	lactam antibiotic by targeting the molecular mechanism of resistance. The	



Tue Nov 23 09:27:47 2004

us-10-822-496-5.rst

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2004, 15:32:53, Search time 2977 Seconds

(without alignments)  
7589.061 Million cell updates/sec

Title: US-10-822-496-5

Sequence: 1 gttcaataatacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST: \*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.6	9.6	997	CNS0057E	AL060767 Drosophila
2	59.4	9.6	987	CNS00418	AL066537 Drosophila
3	57	9.2	536	AG690430	AG690430 nbdx0082E
4	57	9.2	1263	CG745034	CG745034 P037-4-C0
5	56.6	9.1	831	AG301159	AG301159 Mus muscu
6	56.4	9.1	957	CNS00605	AL065951 Drosophila
7	55.8	8.9	1451	AG393206	AG393206 Mus muscu
8	55.4	8.9	1201	CNS0108P	AL098659 Drosophila
9	55.2	8.9	839	AG372305	AG372305 Mus muscu
10	55	8.9	809	AG473719	AG473719 Mus muscu
11	54.8	8.8	1100	CNS0008EY	AL069107 Drosophila
12	54.8	8.8	1209	CG749482	CG749482 P043-3-H1
13	54.6	8.8	793	AG469170	AG469170 Mus muscu
14	54.6	8.8	806	AG344150	AG344150 Mus muscu
15	54.6	8.8	865	AG376220	AG376220 Mus muscu
16	54.6	8.8	959	CNS008WZ	AL052079 Drosophila
17	54.2	8.7	851	AG506735	AG506735 Mus muscu
18	54	8.7	786	AG51082	AG51082 Mus muscu
19	53.8	8.7	391	AG954378	AG954378 nbdx0070D
20	53.8	8.7	901	CNS026A5	AL183110 Tetradon
21	53.4	8.6	1101	CNS00F3N	AL069996 Drosophila
22	53.2	8.6	878	CG926174	CG926174 t086j02ba
23	53	8.5	760	AG534996	AG534996 Mus muscu
24	53	8.5	778	AG474409	AG474409 Mus muscu

25	53	8.5	885	9	CNS018H7	AL109333 Drosophila
26	53	8.5	934	4	BM415736	BM415736 OP20817 M
27	53	8.5	1027	9	AG523385	AG523385 Mus muscu
28	52.8	8.5	1223	9	CNS015ZB	AL106103 Drosophila
29	52.8	8.5	1461	9	AG141873	AG141873 Pan trogl
30	52.8	8.5	1969	3	CG670394	CG670394 Tetradon
31	52.6	8.5	255	2	BE993423	BE993423 Mus muscu
32	52.6	8.5	766	9	AG601460	AG601460 Mus muscu
33	52.4	8.5	756	9	AG429320	AG429320 Mus muscu
34	52.4	8.5	854	9	AG523762	AG523762 Mus muscu
35	52.4	8.5	974	9	CNS049SW	AL280985 Tetradon
36	52.4	8.5	1023	5	BU160131	BU160131 AGENCOURT
37	52.4	8.5	1108	9	CL030161	CL030161 CH216-30C
38	52.4	8.5	1239	9	CG750217	CG750217 P044-3-H0
39	52.4	8.5	1328	9	CG746913	CG746913 P040-2-C0
40	52.2	8.4	756	8	B21789	B21789 F6P1-5P6 IG
41	52.2	8.4	813	9	AG328054	AG328054 Mus muscu
42	52.2	8.4	819	9	AG361489	AG361489 Mus muscu
43	52.2	8.4	1053	4	BM415599	BM415599 OP20676 M
44	52.2	8.4	1101	9	CNS0170C	AL108298 Drosophila
45	52.2	8.4	1157	9	CL076711	CL076711 CH216-152

## ALIGNMENTS

RESULT 1  
CNS0057E 997 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TERT end of BAC #  
DEFINITION BAC12K22 of RPCL-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL060767.1 GI:4943573  
VERSION AL060767  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Bukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 997)

REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

## COMMENT

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuo Osoegawa and  
Aaron Mammosser in Peter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCL-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp; the same strain used for the BDGP's  
PI and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filers for hybridization from the BAC PAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1..997  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone\_lib="BAC12K22"  
/clone\_id="RPCL-98"  
/note="end : TERT"

## ORIGIN

Query Match 9.6%; Score 59.6; DB 9; Length 997;  
Best local similarity 24.7%; Pred. No. 7.4e-05;

Matches	41; Conservative	78; Mismatches	47; Indels	0; Gaps
OY	455	GAATGACACCAAAAGCAATGCGACAGCTGATATGGAGCGCGCGGAGAG	514	
Db	814	RAARRRRAARARARARRRRRRRRARAGRRRRRRRRRRRRRRRRARARAR	755	
OY	515	AGAGAAAGCGCAAAATAGAGGCGAGATTAAGGGGGGAGCGCAAGATTAACGAACT	574	
Db	754	ARAGARAAARRRRRAGAAARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	695	
OY	575	GCAGTGGAGAAAACCTCGAACGAGAAAAAATTTTTTTTTTTTTTTTTTAA	620	
Db	694	RRRRRRGACARRRRGRRRRRGAAGARRRRRRRRRRRRRRRRRRRRRA	649	
RESULT 2				
CNS00418/c				
LOCUS	CNS00418	967 bp	DNA	linear
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR09C16 of RPECI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL066537			
VERSION	AL066537.1			
KEYWORDS	GI:4942778			
SOURCE	GSS.			
ORGANISM	Drosophila melanogaster (fruit fly)			
REFERENCE	Drosophila melanogaster			
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
TITLE	1 (bases 1 to 987)			
JOURNAL	Genoscope.			
COMMENT	Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ; BP 191 91008 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPECI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw ap, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bdgpc.med.buffalo.edu/drosophila_desc.htm">http://bdgpc.med.buffalo.edu/drosophila_desc.htm</a> .			
FEATURES	Location/Qualifiers			
SOURCE	1..987			
	/organism="Drosophila melanogaster"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:7227"			
	/clone="BACR09C16"			
	/clone_lib="RPECI-98"			
	/note="end : TET3"			
ORIGIN				
Query Match	9.6%;	Score 59.4;	DB 9;	Length 987;
Best Local Similarity	32.3%;	Pred. No. 8.4e-05;		
Matches	54; Conservative	61; Mismatches	52; Indels	0; Gaps
OY	454	GGATGAGNACGAAAGACAATGTGGACAGCTGATATGGAGCGCGCGGAACT	513	
Db	878	RGAGRARRRGRRGRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	819	
OY	514	GAGAGAAAGCGCAAAATAGAGGCGAGATTAAGGGGGGAGCGCAAGATTAACGAAAC	573	
Db	818	GGRRRAARRRRRAGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	759	
OY	574	TGCGATGGGAGAAACGTGCAACGAGAAAAAATTTTTTTTTTTTTTTTTTAA	620	

[illegible]

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2004, 15:37:28 ; Search time 90 Seconds

(without alignments)  
4896.546 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620  
Sequence: 1 gtcacaaataacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/5B COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	11.0	7218	1	Sequence 14, Appl
2	52.6	8.5	289	3	Sequence 17, Appl
3	52.6	8.5	289	3	Sequence 3, Appl
4	48	7.7	277	3	Sequence 3, Appl
5	48	7.7	277	3	Sequence 46, Appl
6	43.4	7.0	893	4	Sequence 1, Appl
7	42.6	6.9	123	3	Sequence 1, Appl
8	42.6	6.9	123	3	Sequence 46, Appl
9	42.2	6.8	261	4	Sequence 46, Appl
10	42.2	6.8	1923	4	Sequence 46, Appl
11	41.2	6.6	705	4	Sequence 5061, App
12	41.2	6.6	705	4	Sequence 2043, App
13	40.8	6.6	372	4	Sequence 13, Appl
14	40.8	6.6	372	4	Sequence 13, Appl
15	40.6	6.5	43795	3	Sequence 101, App
16	40.2	6.5	16442	3	Sequence 208, App
17	40.2	6.5	16442	3	Sequence 208, App
18	39.4	6.4	356	4	Sequence 22, Appl
19	39.4	6.4	2327	4	Sequence 22, Appl
20	39.4	6.4	2327	4	Sequence 20, Appl
21	39.4	6.4	2674	4	Sequence 19, Appl
22	39.4	6.4	2771	4	Sequence 18, Appl
23	39.4	6.4	5860	4	Sequence 17, Appl
24	39.4	6.4	9646	3	Sequence 1, Appl
25	39.4	6.4	9646	3	Sequence 1, Appl
26	39.4	6.4	12980	3	Sequence 5, Appl
27	39.4	6.4	12980	3	Sequence 5, Appl

28	39.2	6.3	270	2	Sequence 30, Appl
29	39.2	6.3	270	3	Sequence 30, Appl
30	39.2	6.3	53526	3	Sequence 2, Appl
31	39.2	6.3	53577	4	Sequence 184, App
32	38.8	6.3	548	4	Sequence 894, App
33	38.8	6.3	696	3	Sequence 193, App
34	38.8	6.3	699	3	Sequence 191, App
35	38.8	6.3	717	3	Sequence 189, App
36	38.8	6.3	774	3	Sequence 187, App
37	38.8	6.3	819	3	Sequence 185, App
38	38.6	6.3	1669	3	Sequence 63, Appl
39	38.6	6.2	168174	4	Sequence 2, Appl
40	38.6	6.2	168273	4	Sequence 29, Appl
41	38.4	6.2	260	2	Sequence 29, Appl
42	38.4	6.2	260	3	Sequence 2, Appl
43	38.4	6.2	8642	4	Sequence 18, Appl
44	38.2	6.2	376	2	Sequence 79, Appl
45	38.2	6.2	669	4	Sequence 79, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHIEFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,766  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZapc-Fls  
US-08-232-463-14  
Query Match 11.0%; Score 68; DB 1; Length 7218;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2004, 10:03:27 ; Search time 404 seconds  
(without alignments)  
8056.042 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620  
Sequence: 1 gtcacaaataacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001s:\*  
5: geneseq2002s:\*  
6: geneseq2003s:\*  
7: geneseq2004s:\*  
8: geneseq2005s:\*  
9: geneseq2006s:\*  
10: geneseq2007s:\*  
11: geneseq2008s:\*  
12: geneseq2009s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match length	ID	Description
C 1	47.8	7.7	1840 10	ADC87550 Human GPC
2	47.6	7.7	511 10	ADBS9059 Toxicity
3	47.6	7.7	511 10	ADBS3812 Primary x
4	47.4	7.6	409 4	AA184865 Human pol
5	47.4	7.6	1431 12	ADQ25488 Human sof
6	46.6	7.5	2502 5	AA879317 DNA encod
7	46.2	7.5	210 6	ABL66389 Human ova
8	46.2	7.5	1563 6	ABL53204 Nucleoid
9	46.2	7.5	113033 8	AA154213 SR protei
C 10	46.2	7.4	1426 10	ADC87528 Human GPC
11	45.8	7.4	496 5	AA669552 DNA encod
12	45.8	7.4	4238 10	ADBS3355 Primary x
C 13	45.4	7.3	5059 2	AAK84332 Stealth v
14	45.4	7.3	220895 6	ABK84798 Human cDN
15	45.4	7.3	312477 12	ADPE9744 Human ROC
16	44.6	7.2	654 5	AA66925 DNA encod
17	44.4	7.2	299 4	AA124183 Probe #14
18	44.4	7.2	299 4	ABK63102 Human foe
19	44.4	7.2	299 4	AA149470 Probe #18
20	44.4	7.2	299 4	ABK1297 Human bre
21	44.4	7.2	299 4	ABK16233 Probe #14

22	44.4	7.2	299 4	AAK43403
23	44.4	7.2	299 4	AAK17588
24	44.4	7.2	299 4	ABK43023 Human liv
25	44.4	7.2	299 5	AA109746 Probe #97
26	44.4	7.2	299 6	ABK17496 Human gen
27	44.4	7.2	12700 2	AAV62133 HSV-2 str
28	44.2	7.1	2350 10	ADC87384
C 29	44.2	7.1	2498 2	AAK57416
30	43.4	7.0	893 12	ADJ74215
31	43.4	7.0	893 12	ABV58329 Human pro
32	43.2	7.0	264 5	ABV58274 Human pro
33	43.2	7.0	265 4	AA116963 Probe #68
C 34	43.2	7.0	598 4	AAK16071 Human foe
C 35	43.2	7.0	598 4	AA140968 Probe #96
C 36	43.2	7.0	598 4	ABK28973 Probe #74
C 37	43.2	7.0	598 4	AAK35253 Human bon
C 38	43.2	7.0	598 4	AAK09364 Human liv
C 39	43.2	7.0	598 4	ABK34997 Human liv
C 40	43.2	7.0	598 6	ABK09667 Human gen
C 41	43.2	7.0	598 6	ADK035608 Novel mou
C 42	43.2	7.0	1109 12	ADP81772 Human MD-
C 43	43.2	7.0	67674 12	AA669541 DNA encod
44	43.2	6.9	390 5	AA669541 DNA encod
C 45	42.8	6.9	206 5	ADL44537 Human ova

#### ALIGNMENTS

RESULT 1  
ADC87550/c  
ID ADC87550 standard; DNA; 1840 BP.  
XX  
AC ADC87550:  
XX  
DT 01-JAN-2004 (first entry)  
XX  
DE Human GPCR gene SEQ ID NO:2003.  
XX  
KW db; gene; human; GPCR;  
XX  
KW Ganosine triphosphate-binding protein coupled receptor; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1270724-A2.  
XX  
PD 02-JAN-2003.  
XX  
PF 18-JUN-2002; 2002EP-00013517.  
XX  
PR 18-JUN-2001; 2001JP-00246789.  
XX  
PS 18-JUN-2001; 2001JP-00246789.  
XX  
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
XX  
PI Suwa M, Asei K, Akiyama Y, Aburatani H;  
XX  
DR WPI; 2003-315783/31.  
XX  
DR P-PSDB; ADC87551.  
XX  
PT New polynucleotide, useful for preparing a composition for treating a  
PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.  
XX  
PS Claim 1; SEQ ID NO 2003; 28pp; English.  
XX  
XX The invention relates to a novel polynucleotide encoding a guanosine  
XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
XX the invention may have a use in gene therapy. The polynucleotide and  
XX polypeptide are useful for preparing a composition for treating a patient  
XX in need of increased or suppressed activity or expression of the  
XX guanosine triphosphate-binding protein coupled receptor. The  
XX polynucleotide sequences shown in ADC8548-ADC87616 encode GPCR's of the

CC Invention.

XX Sequence 1840 BP; 58 A; 537 C; 153 G; 768 T; 0 U; 324 Other;

SQ Query Match 7.7%; Score 47.8; DB 10; Length 1840;

XX Best Local Similarity 54.5%; Pred. No. 0.0032; Mismatches 91; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

454 GGAATGAGACCGAAGACATGTGGACAGCTGATATGAGAGCGCGGAGACT 513  
DB GAAAGCGGAAAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 761

514 GAGAGAAAGGCGCAATATGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 573  
DB GAGAGAAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 701

574 TGCACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 620  
DB AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 654

## RESULT 2

ADBS9059  
ID ADBS9059 standard; DNA; 511 BP.

AC ADBS9059;

DT 04-DEC-2003 (first entry)

DE Toxicity-related gene, SEQ ID 4085.

XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;

KW drug screening; toxicity assay; ds.

XX Unidentified.

XX W02003064624-A2.

XX 07-AUG-2003.

XX 31-JAN-2003; 2003WO-US003194.

XX 31-JAN-2002; 2002US-00060087.

XX 15-MAR-2002; 2002US-0364045P.

XX 15-MAR-2002; 2002US-0364055P.

XX 30-DEC-2002; 2002US-0436643P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

XX WPI; 2003-689530/65.

XX Claim 1; SEQ ID NO 4085; 1156bp; English.

CC The present invention relates to a method for predicting a toxic effect  
CC of a tissue or cell sample exposed to the compound, and comparing the  
CC gene expression profile to a database comprising SEQ ID 1-4925, where  
CC differential expression of the gene indicates at least one toxic effect.  
CC The method is useful for predicting at least one toxic effect of a  
CC compound, predicting hepatotoxicity or the progression of a toxic effect  
CC of a compound, identifying an agent that modulates the onset or  
CC progression of a toxic response, predicting the cellular pathways that a  
CC compound modulates in a cell, and identifying an agent that modulates at  
CC least one activity of a protein. The method and compositions of the  
CC present invention using a database of genes having liver toxin-induced  
CC differential expression, are useful in identifying toxicity markers in

CC liver tissues or cells for drug screening and toxicity assays. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 511 BP; 236 A; 46 C; 168 G; 61 T; 0 U; 0 Other;

Query Match 7.7%; Score 47.6; DB 10; Length 511;  
XX Best Local Similarity 55.4%; Pred. No. 0.0021; Mismatches 92; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

455 GAATGAGACCGAAGACATGTGGACAGCTGATATGAGAGCGGCGGAGAGTG 514  
DB GAGAGAAAGGCGCAATATGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 370

515 AGAGAAAGGCGCAATATGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 574  
DB GAGAGAAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 430

575 GCACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 620  
DB AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 476

## RESULT 3

ADBS3812  
ID ADBS3812 standard; DNA; 511 BP.

AC ADBS3812;

DT 04-DEC-2003 (first entry)

DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4354.

XX Toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;

KW toxicity marker; toxicity progression; drug screening;

XX primary rat hepatocyte toxicity modelling; gene; ds.

XX Rattus norvegicus.

XX W02003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

XX 13-MAR-2002; 2002US-0363534P.

XX 08-APR-2002; 2002US-0370248P.

XX 10-APR-2002; 2002US-0371134P.

XX 10-APR-2002; 2002US-0371135P.

XX 11-APR-2002; 2002US-0371150P.

XX 19-APR-2002; 2002US-0373601P.

XX 19-APR-2002; 2002US-0373602P.

XX 22-APR-2002; 2002US-0374139P.

XX 08-MAY-2002; 2002US-0378522P.

XX 09-MAY-2002; 2002US-0378533P.

XX 09-MAY-2002; 2002US-0378655P.

XX 09-JUL-2002; 2002US-0394230P.

XX 09-JUL-2002; 2002US-0394253P.

XX 04-SEP-2002; 2002US-0407688P.

XX 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;  
XX Elashoff M;  
XX WPI; 2003-731472/69.  
XX Determining if a compound induces a toxic effect on a tissue or cell, for

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: November 16, 2004, 13:33:03 ; Search time 36 Seconds  
(without alignments)  
3314.133 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 201  
Sequence: 1 gttcaaaataacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283416 segs, 96216763 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565918

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:  
-MODE=frame\_n2p.model -DEV=xlp  
-O=CGN2.1/USFTC.spool.p/US10822496/runat\_16112004.100157.24608/app\_query.fasta-1.775  
-DB=PIR 75 -QFMT=fastaan -SUFFIX=01112p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=01112p -TRANS=human40.cdi -LIST=45  
-DOCALLGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTWT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10822496.CGN 1.1 63 @runat\_16112004.100157.24608 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6  
-Fgapext=7 -YGAOP=60 -YGAPEXT=60 -DELop=6 -DELEXT=7

Database:

PIR 79: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	9	4.4	106	2 A71615	probable integral
C 2	9	4.4	111	2 S64364	probable membrane
C 3	9	4.5	280	2 T34540	hypothetical prote
C 4	9	4.5	784	2 A33633	transcription repr
C 5	9	4.5	2010	2 B71616	phosphatase (acid
C 6	8	4.0	46	2 T28799	hypothetical prote
C 7	8	3.9	107	2 T29406	hypothetical prote
C 8	8	4.0	127	2 A41061	paired box homolog
C 9	8	4.0	130	2 S65141	hypothetical prote
C 10	8	3.9	139	2 S61040	probable membrane
C 11	8	4.0	167	2 S38112	hypothetical prote
C 12	8	4.0	175	2 T05669	hypothetical prote
C 13	8	4.0	202	2 T11744	dehydrin - kidney
C 14	8	4.0	215	2 I52523	nucleoporin p62 ho

15	8	4.0	244	2 T24631	hypothetical prote
16	8	4.0	246	2 T26265	hypothetical prote
17	8	4.0	246	1 S69770	hypothetical prote
18	8	4.0	249	2 T49239	hypothetical prote
C 19	8	3.9	281	2 J06528	31k major protein,
C 20	8	3.9	282	2 H70006	N-acetylmuramoyl-L
C 21	8	4.0	295	2 S61039	hypothetical prote
C 22	8	3.9	292	2 AB1776	hypothetical prote
C 23	8	4.0	312	2 J05962	paired-box contain
C 24	8	3.9	316	2 S25843	bi-functional cyclo
C 25	8	4.0	323	2 G83291	probable transcrip
C 26	8	4.0	335	2 J05961	paired-box contain
C 27	8	4.0	349	2 J05827	paired-box contain
C 28	8	4.0	349	2 J05202	paired-box contain
C 29	8	4.0	350	2 J05828	paired-box contain
C 30	8	4.0	357	2 E70333	hypothetical prote
C 31	8	3.9	358	2 T15369	hypothetical prote
C 32	8	4.0	362	2 A13288	phosphoribosylamin
C 33	8	4.0	392	2 T09371	hypothetical prote
C 34	8	3.9	401	2 T02958	ribulose-bisphosph
C 35	8	4.0	404	2 A75192	hypothetical prote
C 36	8	4.0	404	2 B71224	hypothetical prote
C 37	8	4.0	419	2 E96501	hypothetical prote
C 38	8	4.0	426	2 T39431	probable methionin
C 39	8	4.0	467	2 T21690	hypothetical prote
C 40	8	4.0	469	2 E43319	carbon-monoxide de
C 41	8	3.9	480	2 S57842	acute myeloid leuk
C 42	8	3.9	535	2 D56680	hypothetical prote
C 43	8	4.0	558	2 G85430	hypothetical prote
C 44	8	4.0	560	2 T05278	hypothetical prote
C 45	8	4.0	561	2 H86442	unknown protein [1

#### ALIGNMENTS

RESULT 1  
A71615  
probable integral membrane protein PFB0415C - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C/Accession: A71615  
R/Gardner, M.J.; Tettein, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
.; Pettes, M.; Salzbey, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;  
Science 282, 1126-1132, 1998  
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A/Reference number: A71600; MUID:99021743; PMID:9804551  
A/Accession: A71615  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-106 <GAR>  
A/Cross-references: UNIPROT:O96177; GB:AE001394; GB:AE001362; NID:93945178; PIDN:AACT187;  
A/Experimental source: clone 3D7  
C/Genetics:  
A/Gene: PFB0415C

Alignment Scores:  
Pred. No.: 1.47  
Score: 9.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 4.41%  
DB: 2  
Gaps: 0

US-10-822-496-5 (1-620) x A71615 (1-106)

QY	620	TTTTTTTTTTTTTTTTTTTTCGCTT	594
DB	5	Phepnepnepnepnepnepneval	13

RESULT 2  
S64364  
probable membrane protein YG069w - Yeast (Saccharomyces cerevisiae)  
N/Alternate names: hypothetical protein G4554

